Washington State Health Care Authority

Health Technology Assessment

Lumbar Fusion (Re-Review)

Draft Evidence Report

August 17, 2015

Health Technology Assessment Program (HTA)
Washington State Health Care Authority
PO Box 42712
Olympia, WA 98504-2712
(360) 725-5126
hta.hca.wa.gov
shtap@hca.wa.gov

DRAFT APPRAISAL DOCUMENT

Lumbar Fusion

August 17, 2015

Daniel A. Ollendorf, PhD	Chief Review Officer
Anne M. Loos, MA	Research Associate
Karin U. Travers, DSc	Research Director
Steven D. Pearson, MD, MSc	President

Table of Contents

List of Acronyms ii

About ICER iii

Acknowledgements iv

Executive Summary 1

ICER Integrated Evidence Ratings 35

1. Background 1

2. Washington State Agency Utilization Data 4

4. Clinical Guidelines and Training Standards 11

5. Medicare and Representative Private Insurer Coverage Policies 13

6. Previous Health Technology Assessments and Systematic Reviews 15

7. Ongoing Clinical Trials 17

8. Methods 19

9. Results 25

9. Recommendations for Future Research 50

References 51

Appendix A: Literature Search Strategy 57

Appendix B: Summary Evidence Tables 59

Appendix C: ICER Integrated Evidence Ratings 73

List of Acronyms

ALIF Anterior Lumbar Interbody Fusion
BMI Body Mass Index
CBT Cognitive Behavioral Therapy
CCI Charlson Comorbidity Index
CI Confidence Interval
CLBP Chronic Low Back Pain
CMS Centers for Medicare & Medicaid Services
DDD Degenerative Disc Disease
FDA Federal Drug Administration
GFS General Function Score
HrQoL Health-related Quality of Life
IRP Intensive Rehabilitation Program
ITT Intent-to-treat
JOA Japanese Orthopedic Association
LCD Local Coverage Determination
MCID Minimal Clinically Important Difference
MCS Mental Component Score
NCD National Coverage Determination
NR Not reported
NRS Numerical Rating Scale
NS Not significant
ODI Oswestry Disability Index
OR Odds Ratio
PCS Physical Component Score
PLIF Posterior Lumbar Interbody Fusion
PLF Posterolateral Fusion
QALY Quality-Adjusted Life Year
RCT Randomized Controlled Trial
RDQ Roland-Morris Disability Questionnaire
RR Risk Ratio
RTW Return to Work
SCL Standard Checklist
SF Short Form
TLIF Transforaminal Lumbar Interbody Fusion
TE Treatment Effect
USD United States Dollars
VAS Visual Analog Scale
WHO World Health Organization
ZDS Zung Depression Scale

About ICER

The Institute for Clinical and Economic Review (ICER) is an independent non-profit health care research organization dedicated to improving the interpretation and application of evidence in the health care system.

There are several features of ICER's focus and methodology that distinguish it from other health care research organizations:

- Commitment to aiding patients, clinicians, and insurers in the application and use of comparative effectiveness information through various implementation avenues, including its core programs, the New England Comparative Effectiveness Public Advisory Council (CEPAC; cepac.icer-review.org) and the California Technology Assessment Forum (CTAF; www.ctaf.org).

- Focus on implementation and evaluation of ICER research to create innovative decision support tools, insurance benefit designs, and clinical/payment policy.

- Deep engagement throughout the process with all stakeholders including patients, clinicians, manufacturers, purchasers, and payers.

- Inclusion of economic modeling in our research, and use of an integrated rating system for comparative clinical effectiveness and comparative value to guide health care decisions.

- ICER's independent mission is funded through a diverse combination of sources; funding is not accepted from manufacturers or private insurers to perform reviews of specific technologies. A full list of funders, as well more information on ICER's mission and policies, can be found at www.icer-review.org.

Acknowledgements

ICER would like to thank the following individuals for their expert opinion as well as peer review of draft documents:

Roger Chou, MD
Professor of Medicine
Division of General Internal Medicine and Geriatrics
Director, Pacific Northwest Evidence-based Practice Center
Oregon Health & Science University

John D. Loeser, MD
Professor Emeritus
Departments of Neurological Surgery and Anesthesia and Pain Medicine
University of Washington

Rajiv K. Sethi, MD
Chair of the Neuroscience Institute
Director of Complex Spine Surgery, Virginia Mason Medical Center
Clinical Associate Professor
Department of Health Services, Program in Health Economics and Outcomes Methodology (PHENOM)
University of Washington

We would also like to thank Erin Lawler at ICER for her contributions to this report.

Executive Summary

Introduction

Low back pain is an exceedingly common complaint and a substantial cause of disability. At any given point in time, more than 10% of individuals are diagnosed with low back pain, and lifetime prevalence ranges from 60-70% in industrialized countries such as the US.[1] The economic impact of low back pain is also substantial. It is the second most common reason for all physician visits in the U.S.[2], and is responsible for approximately $30 billion in direct medical costs annually[3]. In addition, low back pain is associated with substantial indirect costs, in large part due to its detrimental impact on productivity; it is estimated that over 3% of the U.S. work force is compensated for back pain or injury each year[4], with approximately 187 million missed work days and wage losses accounting for an additional $22.4 billion in annual indirect costs[5].

With low back pain often presenting as a temporary condition, and an estimated 25-58% of cases spontaneously resolving[6], nonsurgical, i.e. conservative, treatment is the primary treatment modality at diagnosis. Conservative treatment may include any number of non-surgical therapies, in a structured or unstructured setting, and to lesser or greater degrees of intensity; such therapies include exercise, physical therapy, education, cognitive behavioral therapy, acupuncture, or spinal manipulation. However, persistent low back pain that is refractory to conservative treatment may be seen in as many as one-quarter of patients six months after an initial episode.[7]

Low back pain can be caused by a number of specific and nonspecific conditions, all of which differ in prevalence and affect different age groups. Nerve irritation, muscle strain, and bone or soft tissue damage may all give rise to low back pain. Another common cause of low back pain is lumbar degenerative disc disease (DDD), arising from natural degeneration of an intervertebral disc. DDD is commonly associated with low back pain in many individuals. Use of the term “disease” to describe this condition is something of a misnomer, however, as disc degeneration (dehydration and shrinkage) is a natural consequence of aging, and many individuals never develop overt symptoms of DDD; it is the symptoms of DDD (e.g., pain, limited mobility) that are the primary causes of concern. Diagnosis and subsequent treatment typically involves an initial history and physical examination by a clinician. Depending on the presentation, the clinician might prescribe various conservative self-care therapies or will perform a diagnostic exam to check the patient’s pain tolerance, functional capabilities, and reflexes.[8] An MRI and/or CT scan may be used to identify other potential anatomic causes of the patient’s symptoms, including other co-occurring conditions such as radiculopathy (compression of the root nerve), spondylolisthesis (slippage of a vertebral disc over another, causing spinal instability), or spinal stenosis (narrowing of the spinal canal), lumbar disc herniation (the rupture of an intervertebral disc which then pushes outside its normal boundary).[9,10] The process of disc degeneration appears to be influenced by demographic and behavioral factors (e.g., age, occupation, and activity level), lifestyle (e.g., obesity, smoking), and importantly, genetics.[6]

Multiple treatment options are available for symptoms associated with DDD of the lower back, including conservative measures, minimally-invasive treatments such as spinal injections and radiofrequency ablation, and surgical intervention. Conservative, non-invasive approaches vary widely in method and intensity, and are typically used as a first-line treatment approach for patients complaining of low back pain. When pain becomes chronic (i.e., continues for longer than three months), more intensive

conservative management using interdisciplinary methods is often considered. If these are unsuccessful, management with surgery can be considered. Lumbar fusion surgery, which involves the creation of a permanent connection across the vertebral space by means of a graft, is often considered when conservative treatments fail to relieve the patient's pain.[11] However, many patients may be at risk of continued persistent low back pain, as initial surgery is subject to high rates of reoperation with declining success rates after each consecutive surgery. It is estimated that as many as 80,000 cases of so-called "failed back surgery syndrome" are seen in the U.S. each year, although this figure includes not only fusion but other forms of back surgery.[12]

Not surprisingly, there is significant interest on the part of patients, clinicians, policymakers, and other stakeholders in evaluating the clinical and economic impact of lumbar fusion for patients with chronic low back pain and DDD. An evidence-based inquiry into lumbar fusion as a treatment option for DDD is complicated by the fact that there exists no consensus regarding a true "gold standard" treatment for DDD. Given that lumbar fusion is commonly employed intervention for a number of indications (representing 3.1% of all surgical procedures in the US), a careful evaluation of its effectiveness relative to conservative treatment of DDD will serve to inform policy around its use.

Alternative Treatment Strategies

The major approaches to lumbar spinal fusion and conservative management are described in further detail below. Of note, other minimally-invasive procedures (e.g., spinal injections, denervation procedures) are used in patients with uncomplicated DDD but are not described here given the comparison between surgical and conservative treatment that is the primary focus of our review.

Lumbar Spinal Fusion

During spinal fusion procedures, the spine is stabilized by fusing two or more vertebrae together, using metal rods, bone grafts, or screws.[13] Spinal fusions are classified as either simple (1 or 2 disc levels or a single surgical approach) or complex (more than 2 disc levels or a combined anterior and posterior approach). Fusion may or may not use instrumentation such as screws, plates, or cages. Instrumentation is generally used as an internal splint to hold the vertebrae together while the bone grafts heal. Bone or bone substitutes are used to help fuse the vertebrae together. The bone may be taken from another bone in the patient (autograft) or from a bone bank (allograft). Bone morphogenic proteins may also be used as an alternative to autograft.

During lumbar fusion, the surgeon removes the lamina to help relieve the pressure on the nerve. The surgeon then removes any additional bone that may impinge upon the affected nerve. Bone grafts are then added to the spine; these will eventually fuse with the spine to form a solid bone. Instrumentation may be added to provide additional stability while the grafts heal. There is generally more discomfort experienced after fusion surgery compared to other procedures and recovery takes much longer. Patients usually stay in the hospital for at least three to four days post-procedure. Substantial bone healing takes some time to achieve and the healing process varies from person to person. The indication of bone healing, as evidenced by an X-ray, is not attempted until approximately 6 weeks post-procedure. During this time, the patient's activity must be limited. The surgeon may recommend a post-operative rehabilitation program.[14]

Risks associated with spinal fusion include nerve root damage, bleeding, and infection. While the major risks are relatively rare, the odds of injury are higher with increasing complexity of surgical approach and

use of instrumentation.[15] Other complications, common to all types of major surgery, may include blood clots, myocardial infarction, pulmonary embolism, and pneumonia.
The main approaches to lumbar fusion surgery are as follows:

Posterolateral fusion (PLF)

In a posterolateral fusion, the surgical approach to the spine is from the back through a midline incision that is approximately three inches to six inches long. A bone graft is obtained and laid out in the posterolateral portion of the spine. This region lies on the outside of the spine and is a very vascular area, which is important because the fusion needs blood to supply the nutrients for it to grow. A small extension of the vertebral body in this area (transverse process) is a bone that serves as a muscle attachment site. The large back muscles that attach to the transverse processes are elevated up to create a bed to lay the bone graft on. The back muscles are then laid back over the bone graft, creating tension to hold the bone graft in place. This approach is often considered the "gold standard" for spinal fusion surgery.

Interbody Fusions

Designed to be a less invasive way of obtaining a spinal fusion by using two threaded titanium cylinders to hold the vertebrae in proper position while the spine fusion occurs. These procedures are done using various approaches, and involve removing the disc between two vertebrae and inserting the bone graft into the space created between the vertebral bodies. They are described in detail below:

Posterior lumbar interbody fusion (PLIF)

Unlike the posterolateral fusion, the PLIF achieves spinal fusion in the low back by inserting a cage made of either allograft bone or synthetic material (PEEK or titanium) directly into the disc space. PLIF surgery has a higher potential for a solid fusion rates than posterolateral fusion rates because the bone is inserted into the anterior portion (front) of the spine.

Anterior lumbar interbody fusion (ALIF)

The anterior lumbar interbody fusion is similar to the PLIF approach, except that in the ALIF, the disc space is fused by approaching the spine through the abdomen instead of through the lower back. A three-inch to five-inch incision is made on the left side of the abdomen and the abdominal muscles are retracted to the side.

Transforaminal lumbar interbody fusion (TLIF)

TLIF fuses the anterior (front) and posterior (back) columns of the spine through a single posterior approach. This procedure is done from the back of the spine, differing from the PLIF mainly in the angle at which the disc is approached.

Extreme Lateral Interbody Fusion (XLIF)

An interbody fusion approach in which the surgeon accesses the intervertebral disc space and fuses the lumbar spine using a surgical approach from the side (lateral) rather than from the front (anterior) or the back (posterior).

Conservative, Nonsurgical Management

Conservative, non-invasive approaches vary widely in method and intensity. Further detail on this variability is available in the evidence review. Lower intensity treatments typically include medications, physical and/or exercise therapy, behavioral therapy, chiropractic, and alternative therapy (e.g., acupuncture, yoga). These are typically used as a first-line treatment approach for patients complaining of low back pain. When pain becomes chronic (i.e., continues for longer than three months), interdisciplinary rehabilitation is often considered. Interdisciplinary rehabilitation programs are interventions that combine and coordinate physical, vocational, and behavioral components.[16] These programs are typically physician-directed, with care provided by multiple health care professionals with different clinical backgrounds. The intensity and content of interdisciplinary therapy varies widely; duration of treatment may be as short as one week or as long as 15 weeks and activity levels range from one to eight hours on any given day. Programs typically involve some component of group therapy, usually held in groups of up to 10. Interdisciplinary programs vary not only in duration and intensity, but also in the types of components provided. Worksite interventions, strength training, aerobic exercises, educational interventions, and psychological interventions are all examples of components that can constitute an interdisciplinary program.

Key Questions

The following key questions were felt to be of primary importance for this review:

1. What is the comparative clinical effectiveness of lumbar fusion surgery for patients with chronic low back pain and uncomplicated DDD relative to that of conservative management, minimally-invasive treatments, and other nonsurgical approaches?

2. What are the rates of "treatment success" or "successful clinical outcome" of lumbar fusion as defined by measures of clinically-meaningful improvement in pain, function, quality of life, patient satisfaction, and/or work status?

3. What are the rates of adverse events and other potential harms (perioperative, long-term adverse events, and reoperations) associated with lumbar fusion surgery compared to alternative treatment approaches?

4. What is the differential effectiveness and safety of lumbar fusion according to factors such as age, sex, race or ethnicity, pre-existing conditions (e.g., smoking history), intensity of conservative management (e.g., interdisciplinary rehabilitation vs. physical and/or behavioral therapy alone) technical approach to fusion (e.g., posterolateral vs. interbody, minimally-invasive vs. open procedures), initial vs. repeat surgery, insurance status (e.g., worker's compensation vs. other), and treatment setting (e.g., inpatient vs. ambulatory surgery)?

5. What are the costs and potential cost-effectiveness of lumbar fusion relative to alternative treatment approaches?

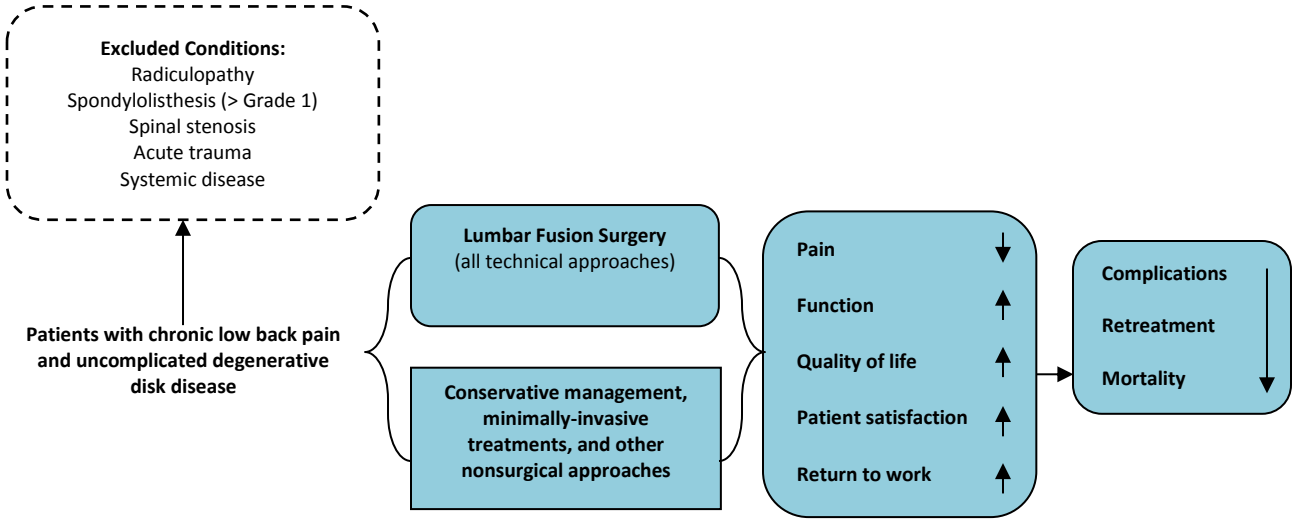
Analytic Framework

The analytic framework for this project is depicted on the following page. We expected that studies would vary substantially in terms of their entry criteria, as there is no agreed-upon standard of what constitutes uncomplicated lower back DDD. In addition, the fusion technique and intensity of the nonsurgical intervention may have differential effects on the outcomes of primary interest in low back

pain studies, including pain, function, quality of life, patient satisfaction, and work status. Finally, randomized control trials (RCTs) of fundamentally different interventions (e.g., surgery for pain relief vs. rehabilitation for functional restoration) may have difficulty enrolling and randomizing patients, resulting in many studies with inadequate statistical power or other quality concerns (e.g., high dropout and/or crossover rates).

There were expected limitations on the available evidence in terms of (a) comprehensive comparisons of lumbar fusion to conservative management, and (b) long-term data on effectiveness and potential harms. As such, judgments about the effectiveness of these interventions rested predominantly upon individual consideration of each type of surgery and its relevant comparators, evaluation of procedure-specific risks, and linkage of shorter-term outcomes to higher-quality data on long-term effects where available.

Figure ES-1. Analytical Framework: Lumbar Fusion



Study Quality

Assessment of the quality of clinical trial reports and systematic reviews followed methods adapted specifically for studies of low back pain from the Cochrane Back Review Group.[17] For observational studies, we used the approach of the U.S. Preventive Services Task Force[18] (see detailed descriptions on the following page). Finally, while there are no published criteria for evaluating quality of case series due to their non-comparative nature, we identified specific quality criteria for inclusion of these studies as follows: (a) sample size ≥100, (b) minimum follow-up of two years, (c) ≥80% patient retention, and (d) ≥75% with uncomplicated DDD or findings stratified by indication for fusion.

***Good**: Meets all criteria: Comparable groups are assembled initially and maintained throughout the study (follow-up at least 80 percent); reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; all important outcomes are considered; and appropriate attention paid to confounders in analysis. In addition, for RCTs, intention to treat analysis is used. Specifically for this review, target or mean/median duration of follow-up did not appreciably differ within study groups.*

***Fair**: Studies will be graded "fair" if any or all of the following problems occur, without the fatal flaws noted in the "poor" category below: Generally comparable groups are assembled initially but some question remains whether some (although not major) differences occurred with follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are addressed. Intention to treat analysis is done for RCTs. Specifically for this review, differences in baseline characteristics and/or duration of follow-up were allowed only if appropriate statistical methods were used to control for these differences (e.g., multiple regression, survival analysis).*

***Poor**: Studies will be graded "poor" if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. For RCTs, intention to treat analysis is lacking.*

Overall strength of evidence for each key question was described as "high", "moderate", or "low", and utilized the evidence domains employed in the AHRQ approach.[18] In keeping with standards set by the Washington HCA, however, assignment of strength of evidence focused primarily on study quality, quantity of available studies, and consistency of findings.

In addition, summary ratings of the comparative clinical effectiveness and comparative value of the procedures of interest (i.e., *across* multiple key questions) were assigned using ICER's integrated evidence rating matrix.[19] The matrix has been employed in previous Washington HCA assessments of virtual colonoscopy, coronary CT angiography, cervical fusion surgery, cardiac nuclear imaging, proton bean therapy, breast imaging in special populations, and bariatric surgery. The matrix can be found in Appendix C to this document.

Of note, our review identified no studies comparing surgery to minimally invasive treatments.

Results

Evidence Quality

The evidence base for comparison of lumbar fusion procedures to non-surgical interventions for uncomplicated DDD has not grown substantially in the past decade, as exemplified by the addition of only one additional RCT[20] since a 2007 review[21] identified four[22-25]. The current review also includes nine cohort studies[26-34], all of which were prospective in nature, with the exception of Smith et al.[34], which was retrospective. The current review also includes three[35-37] case series. A summary evidence table (Table ES-1) capturing the strength of evidence for each of the five key questions of interest can be found starting on page ES-8.

There were a number of specific limitations affecting the quality of the studies in the evidence base. Among these was an imbalance in treatment groups with respect to factors potentially influencing outcomes, or a lack of consideration of such factors in the analysis of the resulting data. Often, but not always, such imbalances were addressed by authors in the analysis phase of the study, presenting treatment effect estimates adjusted for the factors of concern.

Also of concern was the lack of longer-term follow-up data in many studies, and the lack of strict criteria defining treatment groups. Many study populations were subject to substantial attrition rates, limiting

the power of such studies to document effect sizes of interest at these timepoints of interest. Additionally, treatment group definition was often heterogeneous. This precludes easy synthesis of findings with respect to both surgical and non-surgical interventions.
Of the five RCTs identified for this review, we rated three[20,22-24] (60%) to be of good quality based on the comparability of groups with respect to both baseline characteristics and duration of follow-up, and minimized sample attrition; and two RCTs (40%)[20,25] were rated as of fair quality. Quality issues affecting the RCTs are described in detail below. Six[26-29,31,33,34] of the eight prospective cohort studies were rated as good quality (75%), one[32] as fair (12.5%), and one[30] as poor (12.5%). A retrospective cohort study[34] was rated as poor. The poor quality ratings reflect the presence of at least one key quality issue not adequately addressed in either the design or analysis phase of the study.

In the study by Fairbank et al.[24], there was a substantial degree of crossover, with over 25% of patients randomized to intensive conservative management having had surgery by the end of two years; this is in contrast to only 4% of those randomized to surgery who crossed over to conservative management. A separate multiple imputation analysis was conducted to carry forward values for patients who crossed over or were lost to follow-up; this did not materially affect any primary findings. This study also described substantial imbalances between treatment groups in several potentially important baseline characteristics; as with the issue with crossover, the authors addressed this issue in the analysis phase, in this case by estimating the relative treatment effects in multivariate analyses controlling for these factors as additional independent variables.

In contrast to the Fairbank study, crossover rates in either direction between the group randomized to spinal fusion and the group randomized to non-intensive conservative management were relatively low (<10%) in the RCT by Fritzell et al.[25,38], and these crossovers were analyzed separately. However, the authors of this study failed to address any imbalances between the treatment groups with respect to factors possibly impacting treatment outcome; imbalances included mean pain duration between the groups and the presence of comorbidity. An additional limitation of this study included the lack of definition around conservative treatment. These limitations were not severe, but because no effort was made to evaluate their impact, the quality of this study was graded as fair, rather than good.

Two RCTs by Brox et al., were limited by small sample size despite the incorporation of a power calculation in the study design (total sample n=60[22] and n=64[23] in the 2003 and 2006 studies, respectively.) Both studies also had one year of follow-up, somewhat limiting the applicability of the evidence to questions regarding the duration of treatment effect. These limitations were deemed minimal enough to support a quality rating of good for both studies.

The RCT described by Ohtori et al.[20] was also limited by sample size (total sample, n=41), and further by the lack of consistency in the type of fusion surgery performed in the surgical treatment group. These limitations downgraded the quality rating for this study to fair.

Table ES-1: Summary evidence table for lumbar surgery compared to conservative treatment.

Study Information	Comparators	Risk of Bias	Consistency	Directness	Precision	Strength of Evidence	Direction of Effect	Comments
KQ1: Effectiveness of Lumbar Fusion Surgery vs. Conservative Management								
Fusion N=473 RCT=3	Intensive or Interdisciplinary Rehabilitation	Medium	Consistent	Direct	Precise	+++ Moderate	*Comparable* No differences in pain, function, return to work	High crossover rates in some studies
Fusion N=335 RCT=2	Physical Therapy or Exercise alone	Medium	Consistent	Direct	Precise	+++ Moderate	*Comparable* Small benefits seen over 1-2 years of f/u (e.g., faster return to work); differences diminish over time	High crossover rates in some studies
Fusion	Other non- or minimally-Invasive comparators	***NO STUDIES***						

Study Information	Comparators	Risk of Bias	Consistency	Directness	Precision	Strength of Evidence	Direction of Effect	Comments
KQ2: Rates of Treatment Success or Clinically Important Differences								
Fusion N=124 RCT=2	Intensive or Interdisciplinary Rehabilitation	Medium	Consistent	Direct	Imprecise	+++ Moderate	*Comparable* No differences in patient- or observer-rated success rates	
Fusion N=294 RCT=1	Physical Therapy	Medium	Consistent	Direct	Imprecise	++ Low	*Incremental* Higher rates of success or clinical improvement vs. lower-intensity care	Treatment success in 1 RCT, clinically-significant improvement in 1 obs. study
KQ3: Potential Harms of Lumbar Spinal Fusion								
Perioperative Mortality	Harms reported in 14 studies comprising 1,420,986 patients	High	Inconsistent	Indirect	Imprecise	++ Low	Rates of 0.2-0.3% by procedure type	Evidence limited to retrospective databases; most do not isolate DDD
Overall Complications		High	Inconsistent	Indirect	Imprecise	++ Low	Range 9-20% overall Range 1-3% serious	Inconsistent reporting and categorization across studies
Subsequent Treatment	Reoperation or Surgical Revision	High	Inconsistent	Indirect	Imprecise	++ Low	Mean of 12.5% over mean of 5 years of f/u Range 4-32%	Hardware repair, repeat fusion, alternative surgery

Study Information	Comparators	Risk of Bias	Consistency	Directness	Precision	Strength of Evidence	Direction of Effect	Comments
KQ4: Differential Effectiveness and Safety According to Patient, Procedure, or Other Factors								
Intensity of Fusion	Single-level vs. multi-level High vs. low levels of instrumentation	High	Inconsistent	Direct	Imprecise	++ Low	No discernible differences in effectiveness Higher complication rates w/more intensity	Variable estimates by study and procedure
Type of Fusion	Anterior, posterior, transforaminal, combined approaches	High	Inconsistent	Direct	Imprecise	++ Low	Evidence mixed, some studies suggest higher complication rates w/anterior approaches	Variable estimates by study and procedure
Surgical Setting	Inpatient vs. outpatient	***NO STUDIES***						
Conservative Management Intensity	Varying levels of intensity and components	High	Inconsistent	Direct	Imprecise	++ Low	Performance vs. surgery better for more intense programs Evidence mixed for interdisciplinary programs vs. less intense interventions	No discernible patterns of individual program component association with outcome

Study Information	Comparators	Risk of Bias	Consistency	Directness	Precision	Strength of Evidence	Direction of Effect	Comments
Age		High	Inconsistent	Direct	Imprecise	++ Low	Some evidence for greater return to work but also higher disability claims in younger age categories	
Gender		High	Inconsistent	Direct	Imprecise	++ Low	No clear patterns of gender impact	
Race/Ethnicity		***NO STUDIES***						
Workers' Compensation		Medium	Consistent	Direct	Precise	+++ Moderate	Evidence suggesting WC status associated with poorer clinical outcome, lower return to work, and higher costs	WC a predictor in surgical but not non-surgical patients
Psychological Factors		High	Inconsistent	Direct	Imprecise	++ Low	Mixed evidence on effects of depression Presence of neuroses or personality disorder associated with poor surgical outcome	
Lifestyle Factors	Smoking, BMI	High	Consistent	Indirect	Imprecise	++ Low	No association with any surgical outcome of interest	

Study Information	Comparators	Risk of Bias	Consistency	Directness	Precision	Strength of Evidence	Direction of Effect	Comments
KQ5: Costs and Cost-Effectiveness of Lumbar Spinal Fusion								
Surgery	Conservative Mgmt	High	Inconsistent	Direct	Imprecise	++ Low	>$100,000 per QALY over 2 years; other studies had unusual measures or inappropriate comparators	Variable data sources and assumptions; surgical costs high in the US and willingness to pay for fusion lower than for other procedures

Key Question #1: What is the comparative clinical effectiveness of lumbar fusion surgery for patients with chronic low back pain and uncomplicated DDD relative to that of conservative management, minimally-invasive treatments, and other nonsurgical approaches?

We identified three good-quality RCTs, two fair-quality RCTs, four good- or fair-quality longer-term follow-up reports on these RCTs, one fair-quality secondary analysis, one good-quality prospective cohort study, and one poor-quality retrospective cohort study (see Appendix B for study details). Of note, none of these studies compared lumbar fusion to minimally-invasive treatments alone, and conservative management approaches varied across studies. Comparisons are further complicated by differences in study design, methods, and crossover rates. Based on the available evidence, lumbar fusion provides some advantage over lower-intensity conservative approaches (e.g., physical therapy or exercise alone) in improving pain and disability and returning to work over a shorter duration of follow-up (i.e., up to two years); however, differences diminish and are no longer statistically significant over longer durations of follow-up. Conversely, comparisons of lumbar fusion to more intensive and/or interdisciplinary forms of rehabilitation yield no differences in effectiveness.

We identified five RCTs[20,22-25] comparing lumbar fusion to conservative treatment among patients with uncomplicated DDD. Four of these studies[22-25] were evaluated in the original assessment[39] for the HCA; only one additional RCT[20] conducted in Japan was identified for this re-review. Three of these studies[22-24] compared fusion to interdisciplinary rehabilitation with a cognitive-behavioral component. The remaining two RCTs compared fusion to non-intensive physical therapy[25], or an exercise treatment plan[20]. While patients undergoing lumbar fusion had similar absolute levels of improvement in pain and function over one to two years of follow-up across four of the five RCTs[22-25], statistically-significant treatment effects favoring fusion were noted only in the RCTs comparing fusion to less intensive treatment. None of these RCTs included patients who had previously undergone fusion surgery, though three[22,24,25] allowed individuals with who had a prior discectomy.

Table ES-2 on the following page lists the study details of these five key RCTs. Several recent systematic reviews[17,40-42] evaluating these studies have noted that patient inclusion criteria and control treatment regimens may affect outcomes in a substantive way; more details on the effect of the treatment intensity in the conservative cohorts are reported in Key Question #4. The section below describes the short- and longer-term outcomes from these RCTs, as well as the nonrandomized comparative studies we identified as part of our literature search. The rate of harms associated with lumbar fusion versus conservative care are discussed in detail in Key Question #3.

Table ES-2. Study details for 5 key RCTs comparing fusion to conservative treatment in patients with uncomplicated DDD.

Study (Country of Origin)	Quality	Sample size	Setting Type	Entry criteria	Patient characteristics	Control group description	Fusion group description	Follow-up duration
Brox 2003[23] (Norway)	Good	64	Multicenter	Aged 25-60 CLBP ≥1 year *Patients who had undergone previous spinal surgery were excluded*	Age: 43 Pain duration: 10.8 years % male: 39	Cognitive intervention and individual exercises with increasing intensity	Posterolateral fusion with instrumentation and postoperative physiotherapy	1 year
Brox 2006[22] (Norway)	Good	60	Multicenter	Aged 25-60 CLBP ≥1 year *All patients had prior discectomy for disc herniation*	Age: 43 Pain duration: 8.0 years % male: 52 % prior discectomy: 100	Cognitive intervention and individual exercises with increasing intensity	Posterolateral fusion with instrumentation and postoperative physiotherapy	1 year
Fritzell 2001[25] (Sweden)	Fair	294	Multicenter	Aged 25-65 CLPB ≥2 years *Patients with successful discectomy >2 years before fusion were allowed*	Age: 43 Pain duration: 8.0 years % male: 49 % prior discectomy: 18.8	Non-intensive physical therapy + information and education aimed at pain relief	Noninstrumented posterolateral, instrumented posterolateral, or instrumented circumferential	2 years
Fairbank 2005[24] (UK)	Good	349	Multicenter	Aged 18-55 CLPB ≥1 year *Candidates for surgery irrespective of previous root decompression or discectomy*	Age: means reported by age groups Pain duration: 8.0 years % male: 49 % prior discectomy: NR	75 hours of IRP, including daily muscle strengthening and exercise, CBT, and hydrotherapy	At the discretion of the surgeon	2 years
Ohtori 2011[20] (Japan)	Fair	41	Single center	CLPB ≥2 years *Patients who had undergone previous spinal surgery were excluded*	Age: 34 Pain duration: 7.3 years % male: 59	Exercise treatment, including 30 minutes of daily walking and muscle strengthening	Anterior interbody fusion or posterolateral fusion with pedicle screws	2 years

Findings for the key outcomes of interest in available RCTs (i.e., pain, function, and return to work) are summarized below. Further discussion of other outcomes can be found in the full report.

Pain and Function

RCT-based evidence on lumbar fusion surgery versus intensive rehabilitation with a cognitive element comes from three studies[22-24] conducted in Norway and the UK. In the Norwegian RCTs[22,23], no significant differences were observed for pain (as measured by a 100-point VAS scale) or the ODI at 1 year of follow-up; medication use was also not significantly different in either study. Notably, in the later study[22] which included only those patients who had a prior discectomy, absolute changes on the ODI were nominally in favor of the conservative cohort (12.8 vs. 8.9 for surgery). Both studies reported a 97% follow-up rate, with only 2.4% of patients across studies switching to the surgical group after randomization.

Although a significant difference in the ODI favoring lumbar fusion was observed in the UK RCT[24] (-12.5 vs. -8.7, p=0.045) relative to IRP, the authors noted that this difference was only marginally significant. No significant treatment effects were noted for improvements on a shuttle walking test or any of the SF-36 subdomains or component summary scores. These results are confounded by differences between groups for follow-up at two years (78% and 84% in the surgical and conservative groups, respectively), with 28% of patients crossing over to the surgery compared to only 4% switching to the rehabilitation group. However, a separate multiple imputation analysis was conducted to carry forward values for patients who crossed over or were lost to follow-up; this did not materially affect any primary findings.

In the Swedish RCT[25], significant differences favoring surgery were observed in the mean change from baseline to year 2 for both the 100-point VAS (-21.0 vs. -4.3, p=0.0002) and the ODI (-11.6 vs. -2.8, p=0.015) relative to non-intensive physical therapy. However, after six months of treatment the benefits of surgery began to diminish, and the authors observed that back pain increased significantly between one and two years of follow-up for the fusion cohort (p<0.0001). Although this RCT had low attrition with only 2% lost to follow-up, crossover was noted in both groups, including 25% of patients in the rehabilitation cohort and 3% in the surgical group. In the most recent RCT[20] from Japan, there were statistically-significant improvements in favor of ABF and PLF versus exercise treatment (-51.7 and -44.8 vs. -24.0), VAS (-6.1 and -4.0 vs. -3.0), and JOA (+1.4 and +1.3 vs. +0.5) for ABF, PLF, and exercise treatment, respectively, over two years of follow-up (all outcomes, p<0.01). No patients were reported being lost to follow-up, or switching to a different treatment group. However, this small study[20] was largely focused on comparing differences between the two fusion techniques[43], and the control group was only "minimally-treated" with 30 minutes of physician-supervised daily exercises and stretching.

In addition to the above-described RCTs, good-quality follow-up data were available for three of the five RCTs. In a combined study[26] of the original Norwegian RCT cohorts[22,23] (n=124, mean age 43, 45.2% male) after a mean follow-up of four years (with 89% of the original population remaining), the adjusted treatment effect between fusion and non-operative care was non-significant. After nine years[28], patients from both groups (n=99, mean age 43, 38.6% male) who consented to long-term radiography follow-up had similar ODI scores. In a sensitivity analysis which included one-third of patients who crossed over to the surgery group, there were significantly more patients taking opioids on a daily or weekly basis in the surgical cohort compared to non-operated patients (44% vs. 17%, adjusted OR: 4.0; 95% CI: 1.5, 11.0; p=0.005), though no differences were observed in the intent-to-treat analysis. Another fair-quality follow-up study[31] with 261 patients (mean age 42, 47.5% male) pooled from the Brox[22,23] and Fairbank RCTs[24] also found no significant differences between groups on the ODI or VAS, as well as for pain medication use after a mean of 11.4 years of follow-up.

In addition to RCT data, we found one large, good-quality prospective cohort study[33] of 495 patients (mean age 43, 47.5% male) comparing surgery (79% instrumented fusion) to conservative treatment. No specific treatment regimen was prescribed to either patient group in this observational study; rather, patients who were diagnosed with discogenic pain and received surgery within six months were considered part of the surgical group, and all others meeting the inclusion criteria were part of the non-operative cohort. Although the surgical group showed statistically-significant improvements over conventional treatment on the RDQ (-8.8 vs. -1.8) and SF-36 (PCS: +14.5 vs. +2.4) after one year (both outcomes $p<0.001$), the authors noted that the conservative group was minimally-treated, with only 5% receiving CBT, and is likely biased in favor of surgery due to patient selection. Opioid pain medication use was also not statistically-different between groups.

The final study[34] we identified as part of our literature search was a poor-quality retrospective cohort study (n=96, mean age 47, 50% male) comparing lumbar fusion to conservative treatment, which included physical therapy, epidural injections, and medication. This study did not find any significant differences between groups for Numerical Rating Scale pain scores, or the ODI after five years of follow-up. However, there are some substantial methodological concerns with this study, including the failure to control for significant differences in patient characteristics between individuals at baseline and those lost to follow-up, which was more than half of the original population.

Return to Work

Data on the impact of lumbar fusion on return to work come from the Norwegian and Swedish RCTs, and their subsequent follow-up studies. In first Brox study[23], the percentage of employed individuals who returned to work was numerically higher in the intensive rehabilitation control group, but did not reach statistical significance. The 2006 study[22], which evaluated patients with prior disc herniation surgery, similarly found that although there were more patients from the intensive rehabilitation group working full-time, these numbers were too small to be evaluated statistically. In the pooled four-year[26] and 11-year follow-up studies[31], these differences continued to be non-significant.

In contrast, the percentage of patients in the Fritzell RCT[25] not working at baseline due to back pain who were employed at the end of the study was statistically-significantly in favor of the lumbar fusion group (39% vs. 23% for physical therapy, $p=0.049$). The "net" rate of back to work (i.e., subtracting those who stopped working during follow-up) was also significantly higher in the fusion group (36% vs. 13% for physical therapy, $p=0.002$). A subanalysis[29] of the original RCT found that a shorter duration of sick leave prior to treatment was significantly associated with work status at follow-up in both the surgical (14 months for those working, and 31 months for those not working, $p<0.0001$) and conservative (13 months for those working, and 27 months for those not working, $p=0.006$) groups. Other variables, including sociodemographics (e.g., gender, smoking, comorbidity), pain (e.g., duration of pain, quality of pain), clinical findings (e.g., reflexes, sensation), psychological diagnosis (e.g., personality disorders), or radiography (e.g., Modic sign type 1), were not significantly associated with work status at follow-up.

Key Question #2: What are the rates of "treatment success" or "successful clinical outcome" of lumbar fusion as defined by measures of clinically-meaningful improvement in pain, function, quality of life, patient satisfaction, and/or work status?

Much of the work done to quantify clinically-significant improvement in measures of pain and function at the individual patient level came after the publication of the RCTs of interest. Two of the five RCTs we identified for this assessment did not include any measurement of "successful" outcome. Findings from the other three RCTs, as well as one prospective cohort study, mirrored those of continuous measures of effectiveness in that results favoring surgery were limited to studies that compared surgery to minimal or nonspecific approaches to conservative management.

In recent years, multiple efforts have been undertaken to identify clinically-meaningful changes in measures at the individual patient level. These individual "success" outcome measures include a mean 10-20 point change on a 100-point visual analog pain scale or 5-10 points on the RDQ, which are generally considered moderate improvements.[17] Other published thresholds for clinically-meaningful improvement include at least a 30% decrease from baseline on a chronic pain scale or an improvement of at least 20 points on the ODI.[44] Patient-defined minimum acceptable outcomes also include discontinuation of opioid medication and return to some occupational activity, though individuals with significant psychosocial factors (e.g., compensation claims, psychological distress), may be less likely to report satisfaction with treatment despite achieving the desired outcomes.[45]

Unfortunately, the development of measures of clinically-meaningful change at the individual level came after publication of all but the small Ohtori[20] RCT. Measures of treatment success in earlier RCTs were limited to patient-reported or independent observer assessment of improvement after intervention. In the Fritzell RCT[25] comparing fusion to physical therapy of varying intensity, 63% of patients in the surgical group rated their symptoms as "much better" or "better" compared to 29% receiving conservative management ($p<0.0001$). Results were rated as "excellent" or "good" by independent observers for 45% and 18% of patients in the surgical and conservative groups, respectively ($p=0.005$). In contrast, there were no statistically-significant differences in either patient or independent observer ratings of treatment success in the two Brox[22,23] RCTs comparing fusion to cognitive/exercise intervention. Measures of treatment success were not considered in either the Fairbank[24] or the Ohtori[20] RCTs.

Some of the studies include mention of clinically-meaningful change in their Discussion sections. Fairbank[23] and Brox[23] (2003) remark that the mean difference in ODI scores between groups did not approach 10.0, which was considered a clinically-meaningful difference. In fact, the confidence interval in the Fairbank RCT did not include 10.0, essentially ruling out any possible difference in favor of surgery. In the Brox 2006[22] RCT, the observed mean difference in ODI after adjustment for gender and pretreatment expectations was 9.7 points, and the confidence interval around this result included the possibility that exercise/cognitive therapy was *superior* to fusion.

Recent nonrandomized studies have made use of published measures of clinically-meaningful improvement, but their number is extremely limited for patients with uncomplicated DDD. A single good-quality prospective cohort study[33] evaluated clinically-meaningful improvement between treatment groups based on a 30% or 5-point improvement on the Roland-Morris back disability score and found that, after controlling for baseline differences, surgery was significantly better than conservative treatment based on this criteria (57% vs. 25%, $p<0.001$). In addition, 33% and 15% of

patients in the surgical and conservative groups achieved a composite measure of treatment success that included the above Roland-Morris thresholds as well as a ≥30% improvement in pain intensity, no use of opioid pain medication, and a status of employed at 12 months (p<0.001). While these results favored surgery, the authors cautioned that the control group received a variety of interventions and overall, did not appear to receive services consistent with major guidelines for treatment of chronic low back pain. For example, only half of patients received any physical therapy and 5% received a cognitive-behavioral intervention.

Only one case series that met our study inclusion criteria assessed a clinically-meaningful threshold of specific outcome measures for patients undergoing lumbar fusion surgery for uncomplicated DDD. Anderson et al.[46] prospectively evaluated 106 patients who received fusion (ALIF technique with titanium cages and autogenous iliac bone graft) and found that patients who were employed before surgery were significantly more likely to be working after a mean 29.7 months of follow-up (92% vs. 43%, OR 10.5, p=0.0008). An attempt to identify predictors of achieving 30% improvement on the RDQ using multivariate logistic regression found no statistically-significant associations between this outcome and work status, age, smoking history, gender, worker's compensation status, pre-operative pain or RDQ scores, and type of fusion surgery.

Key Question #3: What are the rates of adverse events and other potential harms (perioperative, long-term adverse events, and reoperations) associated with lumbar fusion surgery compared to alternative treatment approaches?

Evidence on harms in published RCTs of treatments for patients with chronic low back pain and uncomplicated DDD is limited by several factors. Many of these studies are too small to capture reliable data on complications that occur infrequently, and the relatively low rate of serious complications has led to standards for research reporting that often do not include a formal assessment of all complications. Other factors contributing to the dearth of data on harms include the lack of observational studies that focus on uncomplicated DDD patients, and the short-term nature of many studies, leading to a failure to observe adverse outcomes associated with surgical interventions that do not manifest until later years (e.g., repeat surgery). Harms associated with conservative treatment are rarely reported and are generally limited to non-compliance with the treatment protocol.

Unlike findings for clinical effectiveness, harms data are often not stratified for interventions that are used for multiple indications (e.g., both uncomplicated DDD and more specific indications). Rather than look to studies comparing different technical approaches of lumbar fusion, which are subject to the same methodological concerns as studies with a non-surgical comparator group (e.g., small sample sizes, shorter duration of follow-up, lack of standardized reporting), we have identified several large database studies evaluating harms associated with lumbar fusion across several indications to provide additional context on the rate of adverse events. These data are evaluated separately from our study set because either the majority of patients did not have a primary indication of uncomplicated DDD, or outcomes were not stratified for this population.

Lumbar Fusion

For lumbar fusion procedures, we have categorized harms as surgery-related mortality, overall adverse events (as reported in the included studies), and requirements for retreatment (e.g., reoperation/revision surgery). Although these studies used various technical approaches to fusion, we

did not make any attempt to stratify outcomes by surgical method. Such data, if available, are summarized for Key Question #4.

Mortality

No data on perioperative mortality attributable to lumbar fusion were reported in any systematic review, RCT, or observational study that met our inclusion criteria. Overall mortality was reported in the Mannion[31] study; 7.1% (10/140) patients died in the fusion group and 0.8% (1/121) patients died in the conservative treatment group during the 11-year follow-up period for the Brox[22,23] and Fairbank[24] cohorts. The authors noted that they could not definitively determine if these deaths were associated with chronic low back pain or its treatment given that some patients had illnesses unrelated to back pain, nor was this difference statistically tested.

Adverse Events

The most frequently-reported adverse events occurred during the perioperative period and included dural tears, bleeding, and wound infection, occurring at a rate of 9-18% in available RCTs and observational studies. Notably, the only RCT published since the original review[20] did not evaluate the rate of complications in either treatment group.

In the Fairbank RCT[24], a total of 19 patients experienced complications from surgery (10.8%), which were primarily dural tears and problems with surgical implants (2.8% each). In the 2003 Brox RCT[23], complications included two wound infections, two bleedings, one dural tear, and one venous thrombosis. Overall 6 patients (18.2%) experienced a complication, and all presented as early complications; there were no late complications associated with surgery. The 2006 Brox RCT[22] reported wound complications in only two patients (8.7%). During long-term follow-up for these studies[26], no additional complications related to surgery were reported. Fritzell et al.[25] reported 53 early complications occurring in 17% of patients, and 13 (6%) of patients suffered a late complication (defined as more than two weeks after surgery), including 9 patients who developed nerve root pain related to the pedicle screw implant. Overall, there were 16 (7.8%) unintended reoperations related to compilations in the fusion cohort.

We identified only one small, poor-quality prospective comparative cohort study[30] which evaluated outcomes for patients with degenerative disc disease (n=46, mean age 55, 59% male) undergoing minimally-invasive fusion surgery compared to those who had a previous discectomy undergoing fusion for the first time. Although more patients in the revision group experienced dural tears, overall there were no statistically-significant differences in perioperative complications between the groups.

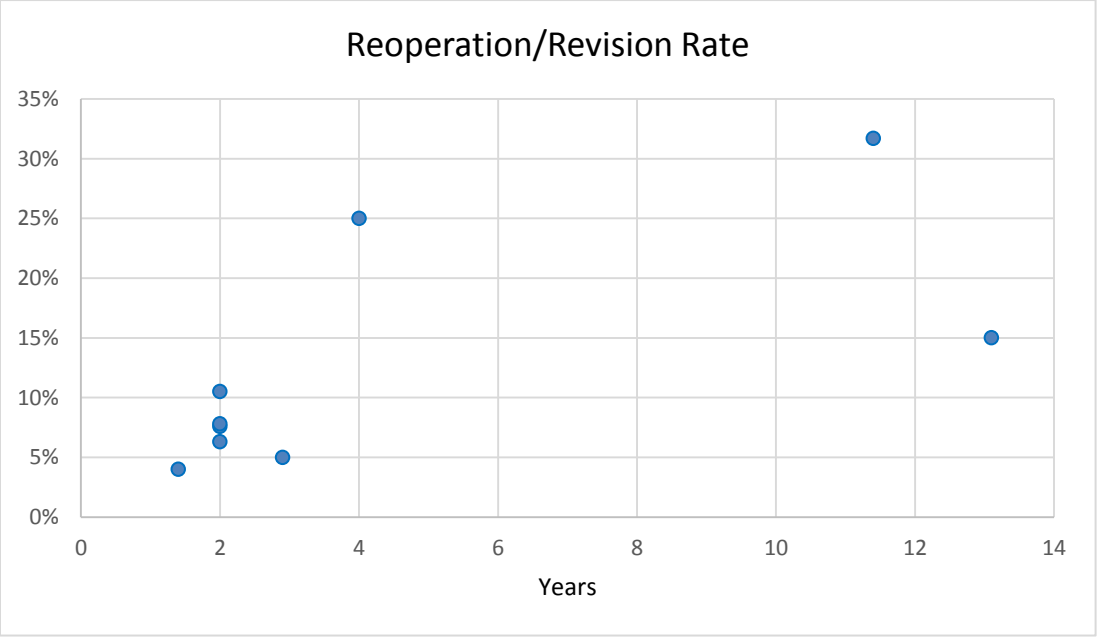
One case series[36] of 118 patients did not reported any intraoperative or major complications after surgery, but 2 patients (<1%) experienced a hematoma and one patient received a permanent disability rating. Complications rates in case series tend to be lower than in RCTs and cohort studies, which is not surprising given the information biases attendant in evaluations.

Subsequent Treatment

Data from available studies indicate that requirements for additional surgery vary widely in both reported rate and indication for such surgery. Across all studies, the rate of reoperation and/or revision surgery averaged approximately 12.5% across studies over a mean of five years of follow-up. As shown in Figure ES-2 on the following page, reoperation continues to be a concern even years after initial surgery. Studies of shorter duration (i.e., up to two years) had a lower reported rate of reoperation (4%-11%) compared to the limited number of studies with longer follow-up periods (15%-32%). Indications

for additional surgery include hardware removal, repeat fusion, alternative lumbar surgery (e.g., discectomy), or some combination. The figure below represents those studies[24-26,28,30,31,33,36,37] in our set that reported on the rate of reoperations. It is difficult to distinguish between revision surgery and reoperations for two reasons: 1) studies often use these terms interchangeably, and 2) patients can undergo surgery for multiple indications (e.g., a combination of hardware removal and repeat fusion), so reasons for repeat surgery are not always stratified. One study[31] reported these outcomes separately; of the 38 (15%) patients requiring additional surgery, 17 involved hardware removal, 11 required repeat fusion, nine had a combination procedure, and one underwent a discectomy.

Figure ES-2. Rate of reoperations/revision procedures across all studies reporting this outcome.



Interestingly, only one[36] of these studies associated repeat surgery with adjacent segment degeneration, which is considered a major concern with lumbar fusion[47] and can cause recurrent lumbar pain. Lammli and colleagues reported that one-third of the additional surgical procedures were performed due to degeneration adjacent to the primary fusion level.[36] Two additional long-term studies in our sample evaluated this outcome but with conflicting results. Froholdt et al.[27] (n=48, mean age 43, 42.8% male) included patients from the Brox RCTs[22,23] who had radiographs available for review, and found no differences between the surgical and conservative groups after a mean of nine years follow-up. In contrast, another follow-up study[32] which included 369 patients (mean age 43, 46.7% male) who participated in the Brox[22,23], Fairbank[24], and Fritzell[25] RCTs who consented to long-term radiographic follow-up over a mean duration of 13.1 years found a significant correlation between surgery and adjacent segment degeneration by assessing adjacent disc height (TE: -0.44 standard deviations; 95% CI: -0.77, -0.11; p=0.01), but this correlation was not associated with statistically-significant changes in patient-reported measurements of pain or disability.

Conservative Care

No data on observed mortality or complications due to conservative treatment were found in the available evidence. Information was limited to rates of crossover to surgery as summarized above.

Large Database Studies of Lumbar Fusion

As mentioned previously, we identified six large database studies evaluating complications for fusion across several indications (e.g., stenosis, isthmic spondylolisthesis, scoliosis, etc.) that did not meet our inclusion criteria but are described here to provide additional information on complications associated with lumbar fusion. Three studies[48,49] used the National Inpatient Sample (NIS) database, two[50,51] studies evaluated data from Washington State-specific databases, and one study[52] reviewed the Swedish Spine registry for the 2011 calendar year.

The most recent study[49] to use NIS data to evaluate three different primary interbody fusion cohorts (923,038 fusions) over nine years. Patients with uncomplicated DDD represented a majority of patients for each fusion group (80.1%, 60.6%, and 78.6% for anterior lumbar interbody fusion [ALIF], posterior/transforaminal lumbar interbody fusion [P/TLIF], and combined anterior-posterior interbody fusion [APF], respectively, mean across groups: 64.2%). Table ES-3 on the following page represents the rate of complications among these groups, showing a significantly higher rate for APF for 12 of 16 complications, and a significantly higher rate of mortality for ALIF. These rates were not adjusted for differences in baseline characteristics.

Those studies[48,53] that did not meet our inclusion criteria (primarily because they did not have a majority of patients with uncomplicated DDD or report outcomes specific to this population), but reviewed large samples from the NIS database, evaluated whether mortality was associated with the incidence of specific complications of lumbar fusion across multiple diagnoses. The first study[48] identified a sample of 220,522 patients who had a fusion procedure (ALF, PLF, or APLF) for degenerative diseases of the lumbar spine and found that the incidence of postoperative ileus was significantly higher in those who had ALF surgery relative to PLF surgery (4.9 vs. 26.0 per 1,000) . Presence of postoperative illeus was associated with significantly higher Charlson comorbidity index (CCI) scores (3.05 and 2.13 for PLF and ALPF, respectively, $p<0.001$), and rates of mortality in both the ALF (1.5 vs. 4.1 deaths per 1,000, $p=0.025$) and PLF (1.1 vs. 4.0 deaths per 1,000, $p<0.001$) fusion groups. The second study[53] evaluted the incidence and potential risk factors of cerebral vascular accidents (CVA) following lumbar fusion surgery. A total of 340 CVAs out of 264,891 fusions (1.3 per 1,000) were identified between 2002-2011, and were associated with a greater mortality rate (73.7 vs. 0.8 per 1,000 patients) compared to those who did not have a CVA. Risk factors associated with CVA include advanced age (64.4 vs. 55.0 years for no CVA) and preoperative comorbidies as demonstrated on the CCI (4.03 vs. 2.52 for no CVA) (both outcomes, $p<0.001$).

Table ES-3. Comparison of complications among P/TLIF, ALIF, and APF[49].

Complications	ALIF (%)	P/TLIF (%)	APF (%)	p-Value
Mortality	**0.25**	0.15	0.18	<0.001
Dysphagia	**0.17**	0.13	0.11	0.0017
Device Related	**5.43**	2.44	3.89	<0.001
Neurologic	0.37	**0.96**	0.55	<0.001
Cardiac	0.90	0.87	**1.23**	<0.001
Peripheral Vascular	0.22	0.08	**0.28**	<0.001
Respiratory	1.65	1.25	**2.20**	<0.001
Gastrointestinal	4.83	2.20	**5.56**	<0.001
Genitourinary	0.84	1.02	**1.11**	<0.001
Postoperative Shock	0.08	0.08	**0.13**	0.0002
Hematoma/Seroma	0.63	0.62	**0.82**	<0.001
Intraoperative Accidental Puncture/ Laceration of Nerve/Blood Vessel	3.41	3.43	**4.20**	<0.001
Wound Dehiscence	0.27	0.15	**0.37**	<0.001
Postop Infection	**0.74**	0.43	**0.74**	<0.001
Acute Anemia Secondary to Hemorrhage	7.39	11.42	**11.60**	<0.001
Acute Respiratory Distress Syndrome	**1.36**	0.75	**1.36**	<0.001
Venous Thromboembolic Events	0.62	0.41	**0.73**	<0.001

Table key: ALIF, anterior lumbar interbody fusion; APF, anterior-posterior interbody fusion; ARDS, acute respiratory distress syndrome; CNS, central nervous system; GI, gastrointestinal; GU, genitourinary; P/TLIF, posterior/ transforaminal lumbar interbody fusion; VTEs, venous thromboembolic events.

Note: Highest percentage is given in **bold**. p Value is from chi-square test.

Two additional dataebase studies[50,51] reviewed Washington state-specific data to identify complications and mortality associated with lumbar fusion procedues. One of these studies[50] used the Comprehensive Hospital Abstract Reporting System (CHARS) registry of all nonfederal hospitals in Washington State and identified 5,091 adults who underwent a primary fusion procedure for degenerative diseases of the lumbar spine between 2004 to 2007. The overall complication rate for patients with DDD (n=1,097 or 18% of the total population) within the first 90 days after surgery was 4.2%, 2.1% had a repeat lumbar fusion surgery, and there were no deaths. During the one year follow-up, an additional 3.2% had a reoperation, but no deaths or complications were observed. The second study[51] identified all workers' compensation claimants (n=2,378) who underwent fusion from 1994 through 2001 and found a 90-day perioperative mortality rate of 0.29% (95% CI, 0.11%, 0.60%) and a 3-year cumulative mortality rate of 1.93% (95% CI: 1.41, 2.57). Interestingly, patients without a specific indication for surgery were more likely to experience the adverse consequences of narcotic use; a diagnosis of degenerative disc disease was associated with the highest risk of analgesic-related mortality (Risk Ratio [RR] 2.71; 95% CI: 1.17, 6.28).

The final database study[52] retrospectively reviewed the Swedish National Spine Register from 2011. In a cohort of 3,066 patients who had fusion surgery, 14% underwent reoperations over a mean three years of follow-up, of which 53% were related to removal of an implant and 47% were related to other complications from surgery. A minority of patients (8%) were listed as having a sole diagnosis of DDD and 38% of patients had previous lumbar spinal surgery; however, no further details on complications were reported.

Key Question #4: What is the differential effectiveness and safety of lumbar fusion according to factors such as age, sex, race or ethnicity, pre-existing conditions (e.g., smoking history), intensity of conservative management (e.g., interdisciplinary rehabilitation vs. physical and/or behavioral therapy alone) technical approach to fusion (e.g., posterolateral vs. interbody, minimally-invasive vs. open procedures), initial vs. repeat surgery, insurance status (e.g., worker's compensation vs. other), and treatment setting (e.g., inpatient vs. ambulatory surgery)?

There is little evidence to suggest that greater surgical intensity is related to changes in outcome in the long-term; advantages to less-intensive surgery (e.g., the effect of minimally-invasive surgery compared to open surgery was positive on HRQoL[18]) were noted in the short-term but did not persist in longer-term follow-up >2 years. On the other hand, our review suggests that more intensive and interdisciplinary rehabilitation featuring behavioral intervention may be both superior to usual-care approaches featuring only physical or exercise therapy, and that these more intensive approaches produces comparable outcomes compared to lumbar fusion. Workers' compensation status appears to have a differential treatment impact, negatively affecting some surgical outcomes (but not those of conservative management). This impact on surgical outcomes was inconsistent, however, as were the impact of age and gender. Our review did not find smoking status or BMI to be predictive of surgical outcome. These findings suggest that it will be difficult to use such factors to define subgroups of patients with DDD in whom surgical or conservative interventions would be preferentially indicated.

There are scant and often conflicting data addressing intervention-associated and patient-based factors that may influence outcomes following treatment for uncomplicated DDD. Several factors (e.g., age, gender, complexity of fusion) are often adjusted for in analysis of the effect of treatment for DDD on various outcomes of interest; however, the rationale for variable selection and/or results of stratified analyses suggesting differential effects are rarely provided.

The evidence on differential effects of lumbar fusion according to various patient- and treatment-defined subgroups is summarized in the sections that follow. The focus of attention in this executive summary is on factors with evidence of material impact on outcome; a full treatment of all factors assessed is available in the full report.

Surgical Intensity

Within the primary review scope, patients undergoing spinal fusion had similar levels of improvement in pain and function over one to two years of follow-up across all five identified RCTs comparing surgical to conservative treatment. Statistically-significant treatment effects favoring fusion were noted only in the RCTs comparing non-intensive physical therapy or exercise to PLIF without decompression or ALIF or PLIF with or without variable screw placement.[25] This is in contrast to a lack of significant findings in RCTs comparing intensive conservative management strategies to PLIF with posterior transpedicular screws[25] or to a range of fusion options[24].

Our review did not identify any publications describing the impact of previous surgery on the relative effect of surgical intervention for uncomplicated DDD compared to intensive conservative therapy. Two RCTs reported no benefit of lumbar fusion over intensive conservative management among patients with previous surgery for disc herniation[22,23]; this finding mirrors the lack of benefit noted for lumbar fusion over intensive conservative management among patients with no previous surgery Additionally, a prospective study of minimally invasive TLIF performed in 25 patients as a primary surgical intervention and in 21 patients as a revision documented no pain or function differences between primary and revision surgery at 1 year; these findings support the observation that there are few differences in primary versus revision surgery among patients with uncomplicated DDD treated with a surgical intervention.

The impact of the level of fusion on relative treatment effect of fusion versus conservative management was not evaluated in the five RCTs identified in our review. Our review also identified one case series describing outcomes in a population of 106 patients with discogenic back pain followed for a mean of 29.7 months after treatment with varying intensity of ALIF (according to level of arthrodesis).[46] Using a multivariate regression model, the authors evaluated the impact of single versus multiple-level fusion on a number of different outcome measures: return to work, a 30% improvement in the VAS pain score, or an increase of at least 30% on the Roland Morris score.[46] Fusion level was not found to be statistically-significantly associated with any of these outcomes.[46] Outside the body of primary literature identified within the scope of this review, several reports offer additional information regarding the impact of lumbar fusion of varying intensity. The impact of differing levels of fusion (1, 2, or 3 or more) was evaluated in a retrospective study of 143 active duty military personnel (mean age 36.3 years), of whom 118 (83%) were DDD patients, treated with TLIF and followed for a mean of 34.9 months. [37] The level of fusion was not associated with the likelihood medical separation (i.e., an inability to remain on active duty).[37]

The impact of minimally-invasive vs open surgery was evaluated in a systematic review reporting the impact of these two approaches to PLIF surgery.[18] The findings of this review suggest that minimally invasive techniques may be associated with better HRQoL outcomes in the short-term, though the effect was variable, and not present at all in longer term follow-up (>2 years).[18] Primary reports reflecting these findings include a prospective study of 66 patients undergoing single level TLIF, comparing those experiencing open (n=33, of which 14 were patients with DDD) versus minimally invasive surgery (n=33, of which 13 were patients with DDD),[54] there were significantly lower VAS pain scores at 6 months post-surgery among those treated with a minimally invasive approach; no longer term data were presented.[54] Likewise, retrospective study of 64 patients receiving either minimally invasive TLIF or open TLIF for the treatment of DDD or spondylolisthesis reported lower VAS pain scores in the early post-operative period for the minimally invasive treatment, with no longer term data presented.[55]

The impact of instrumentation in lumbar fusion surgery was evaluated in a retrospective analysis of 1,310 DDD patients undergoing lumbar fusion, examining the impact of varying levels of surgical instrumentation on HRQoL, pain and function, and return to work.[56] Patients undergoing non-instrumented fusion (n=115) had higher levels of pain as measured on a VAS scale than those undergoing instrumented interbody fusion (p=0.02), although no differences in either HRQoL (as measured using the EQ-5D) or disability (as measured using the ODI) were noted.[56] Another randomized trial of patients with DDD treated with PLF (n=72) vs PLIF (n=73) reported no ODI or VAS differences between the 2 groups at 36 months.[57] These findings were supported by a prospective study of patients with DDD treated with PLF (n=82) and PLIF (n=80), in which no difference between the 2 groups was noted for ODI.[58]

The impact of cage use in lumbar fusion surgery was evaluated in a recent systematic literature review, which reported that single cage lumbar interbody fusion had significantly lower rates of complications than did two-cage fusion surgery (OR 0.30 [0.10, 0.95]).[59] Supporting this finding are those of a retrospective population-based cohort study of 2,378 chronic back pain patients treated with lumbar fusion surgery and receiving Washington State workers compensation who were followed by for a mean of 6.6 years; in this study, the use of cages or instrumentation was associated with increased complication rate compared with bone-only fusion surgery (OR 2.20 [1.16, 4.16]), without any improvement in disability or reoperation rates.[60]

Surgical Approach

The primary focus of our review was on comparisons of lumbar fusion to non-operative management; we nevertheless summarize available data comparing different forms of fusion below, with a focus on uncomplicated DDD where possible.

There exists little conclusive evidence documenting the impact of surgical approach on the outcomes of lumbar fusion among patients with uncomplicated DDD. A five-year RCT comparing the clinical outcomes of posterior midline fusion (n=25) compared to a paraspinal approach (n=25) in DDD patients reported significant improvement in outcomes for both groups, but no differences between groups.[61] Another RCT with 2 years of follow-up reported no statistically significant differences in function (ODI) or pain (VAS) between groups of DDD patients with radiculopathy treated with TLIF (n=51) and PLF (n=47).[62] Evaluating the hypothesis that APF, with its anterior approach, may result in a higher incidence of major complications than TLIF; a respective analysis of 68 DDD patients treated with APF compared to 65 with TLIF reported higher rates of intra-operative complications associated with APF, and higher rates of post-operative complications associated with TLIF, with similar clinical outcomes in both groups.[63] A retrospective database analysis similarly documented a significantly increased incidence of postoperative ileus ALF surgery compared to PLF surgery (74.9 vs. 26.0 per 1,000; $p<0.001$).[48] A prospective observational study documented two-year outcomes associated with posterior fusion with translaminar screw fixation compared to TLIF in a cohort of 120 patients with DDD, and reported no difference in either clinical outcomes or treatment satisfaction.[64]

Conservative Management Intensity

Conservative management in the five identified RCTs incorporated a range of options, and differed in intensity. Table ES-4 on the following page describes the various components of the conservative management programs in each of the RCTs identified.

The conservative management programs differ with respect to the intensity of the intervention, with three[22-24] programs providing intensive treatment over a period of less than one month, and another two[20,25] providing treatment either over a longer period of time or with an undefined intensity. While comparisons across these RCTs are complicated by differences in study design, methods, and crossover[40], there are discernable patterns. Patients undergoing spinal fusion had similar levels of improvement in pain and function over one to two years of follow-up across all five identified RCTs comparing surgical to conservative treatment outcomes. However, statistically-significant treatment effects favoring fusion were noted only in the two RCTs comparing fusion to non-intensive physical therapy or exercise.[20,25] However, there appears to be relative benefit conferred by intensive non-surgical management compared to surgery.[22-24] No particular component of the management programs appears to be substantially associated with a greater relative benefit compared to surgery; such greater

relative benefit appears instead associated with structure and intensity of the program over the short-term perioperative period.

Table ES-4. Components of Conservative Management Programs Incorporated as Comparators in RCTs Evaluating Lumbar Fusion in the Treatment of Uncomplicated DDD

	Conservative Management Components						
Publication	**Comparator**	**Strength Training**	**Aerobic Exercise**	**Educational Interventions**	**Biopsychosocial Interventions**	**Other Interventions**	**Program Intensity**
Brox, 2003[23]	PLIF		Individualized endurance and coordination exercises	Rehab specialist lecture -Daily reinforcement	Fear avoidance Belief modification		75 hours/ 3 weeks
Brox, 2006[22]	PLIF			Rehab specialist lecture -Daily reinforcement	Fear avoidance Belief modification		75 hours/ 3 weeks
Fairbank, 2005[24]	Various fusion	Muscle stretching Spinal flexibility General strength Spine stability	Individualized endurance and coordination exercises		CBT: Fear avoidance and belief modification	Hydrotherapy	60-110 hours/ 3 weeks
Fritzell, 2001[25]	PLF, ALIF, or PLIF*	Ad hoc physical therapy		Ad hoc educational programs	Ad hoc cognitive training		NR
Ohtori, 2011[20]	ALIF or PLIF*	½ hour daily muscle stretching	1 hour daily walking				1095 hours/ 2 years

*Statistically significant treatment effect of surgery over conservative management

Our review did not identify any studies directly comparing conservative management programs of varying intensity. Outside the scope of our review, there is evidence describing the relative effectiveness of varying intensity of conservative management. Several RCTs describe the efficacy of intensive interdisciplinary rehabilitation programs compared to specific physical therapy regimens.[65,66] Findings from those RCTs comparing higher intensity conservative management to some form of physical therapy were consistent, in that no significant treatment effects favoring the more intensive program were observed for any primary outcome measure; substantial improvements in pain, disability, and function were observed in both treatment groups.[65,66] Several systematic reviews describing the effectiveness of higher intensity programs have also been published. One review found that intensive interdisciplinary rehabilitation programs (>100 hours) were associated with clinically-important improvement in function[67] compared to usual care, while another did not find such an association between program intensity and clinical benefit[68]. In sum, there is moderate evidence that intensive conservative management programs confer some level of incremental benefit over usual care, but not necessarily over less intensive programs of physical therapy.

Sociodemographic Factors

Age

Our review identified three good quality studies evaluating age as a potential predictor of treatment outcome: one RCT[29] and two case series[46,69]. The RCT [29] is a secondary analysis of data derived from the Swedish Lumbar Spine Study as described above[25,38]. The authors found that working status at the end of the 2-year follow-up was associated with younger age (evaluated as a continuous variable) in the surgical treatment group, but not in the non-surgical group, indicating a differential impact of age on treatment.[29] Supporting the impact of age on return to work was another case-series identified by our review, of 620 patients with DDD treated with single level posterolateral fusion, followed at least 3 years, of whom 24.4% returned to work in within 2 years postoperatively.[35] Negative predictors of return to work included age more than 50 years at fusion (OR 0.66; 95% CI: 0.45, 0.95).

Our review also identified another good quality case-series with results contrasting with those above. This study described outcomes in a population of 106 patients with discogenic back pain treated with ALIF and followed for a mean of 29.7 months.[46] Using a multivariate regression model, the authors evaluated the impact of age on a number of different outcome measures: returning to work, a 30% improvement in the VAS pain score, or an increase of at least 30% on the Roland Morris score.[46] Age was not found to be associated with any of these outcomes.

Outside of the scope of the current review, there are conflicting data around the relationship between age and the outcome of surgical treatment for uncomplicated DDD. A retrospective population-based cohort study of 2,378 chronic back pain patients treated with lumbar fusion surgery and receiving Washington State workers compensation followed by for a mean of 6.6 years reported that age greater than 30 was significantly associated with higher rates of work disability, to the greatest degree in the oldest age group, greater than 60 (OR 3.07 [1.71-5.51]) compared to the reference group (those below 30). In contrast to this finding, another retrospective study, of 143 active duty military personnel (mean age 36.3 years), of whom 118 (83%) were DDD patients, treated with TLIF and followed for a mean of 34.9 months, younger age was associated with medical separation (an inability to remain on active duty) (OR for each additional year of age 0.93 [0.87, 0.98], p=0.01).[37]

Workers' Compensation

Our review identified one good quality RCT, and one good quality case-series describing workers compensation as a potential predictor of the impact of DDD treatment. The RCT[29] is a secondary analysis of data derived from the Swedish Lumbar Spine Study[25,38]. The authors found that WC status was negatively associated with patient global assessment (p=0.049) and work status (p=0.035) in the surgical group, but not in the non-surgical group, indicating a differential impact of WC on treatment[29].

Our review also identified one good quality case-series describing outcomes in a population of 106 patients with discogenic back pain treated with ALIF and followed for a mean of 29.7 months.[46] Using a multivariate regression model, the authors evaluated the impact of WC on a number of different outcome measures: returning to work, a 30% improvement in the VAS pain score, or an increase of at least 30% on the Roland Morris score (measuring function).[46] WC status was not found to be statistically-significantly associated with any of these outcomes. The multivariate model also included pre-surgery work status as a potential predictor of outcome, and this was independently associated with return to work (OR 10.5 [2.64, 41.4], p=0.0008), but not with VAS pain score or Roland Morris function score.

Outside of the scope of this review are several sources of information which may further illustrate the variation in findings around the impact of WC status on the outcome of treatment of back pain patients. In contrast with the inconsistent findings above, compensation status, whether through litigation or workers' compensation, is in general consistently associated with poor outcomes after any surgical intervention, as reported in a systematic review of 211 clinical trials with relevant information.[70] Several relevant publications describe primary studies of lumbar fusion which add additional specific evidence to the association of WC and outcomes in groups treated thusly. A nonrandomized comparative prospective study of 66 patients undergoing single level TLIF compared those experiencing open (n=33, of which 14 were patients with DDD) vs minimally invasive surgery (n=33, of which 13 were patients with DDD).[54] This study found no significant differences in clinical outcomes between those receiving WC compared to the non-WC group, either overall, or stratified by the open versus minimally invasive technique.[54] These findings were in contrast to those of a prospective non-comparative study of 125 patients undergoing ALIF over a 2-year period (of whom 27 were patients with uncomplicated DDD), which documented a significantly lower rate of clinical success (as defined by a score of 1 or 2 on the PSI) among patients receiving WC (68% success rate) compared to those not (91% success rate) (p=0.006).[71] This negative relationship did not hold true in the analysis of either the ODI or the SF-12 Physical Component Summary (PCS) or Mental Component Summary (MCS).[71]

Psychological Factors

Our review identified two good quality studies describing psychosocial factors as potential predictors of the impact of treatment. The first was performed in the context of a good quality multicenter RCT.[29] This study[29] is a secondary analysis of data derived from the Swedish Lumbar Spine Study[25,38]. In this analysis of data from 294 enrolled patients, the authors evaluated factors they deemed as potential predictors of various treatment outcomes in surgical and conservative (non-intensive physical therapy) patient groups.[29] Outcome measures included reduction of disability (≥50% reduction of the ODI score), patient global assessment of treatment effect (improvement/no improvement), and work status at the conclusion of 2 years of followup.[29] Using a stepwise, forward multiple logistic regression analysis, the authors found that neurotic personality (measured using the Karolinska Scales of Personality) was statistically-significantly negatively associated with improvement in patient global assessment in the

surgical group (p=0.006).[29] However, this association was not significant in the non-surgical group, indicating a differential impact of neurotic personality traits on treatment.

Conversely, in this same study, depressive symptoms measured using the Zung Depression Scale were negatively associated with improvement in the patient global assessment score in the conservative group but not in the surgical group, suggesting as well a differential impact of this trait on treatment.[29] There was no association, differential or otherwise, noted between depression and either ODI or work status in either the surgical or non-surgical treatment groups.[29]

Our review also identified a good quality retrospective case series of 620 patients with DDD treated with single level posterolateral fusion, followed for at least 3 years, of whom 24.4% returned to work in within 2 years postoperatively.[35] Negative predictors of return to work included psychological comorbidity (defined as undergoing psychotherapy) before fusion (OR 0.30; 95% CI: 0.14, 0.62).[35]

Outside of the scope of the current review, there are data which may further illustrate nuances of the relationship between psychological comorbidities and outcomes of treatment for uncomplicated DDD. A systematic literature review documented that psychological factors may in fact modify the treatment effect of fusion versus conservative treatment, with the outcome of fusion less favorable among patients with personality disorder, neuroticism, or depression.[72] Supporting these findings is a retrospective population-based cohort study of 2,378 chronic back pain patients treated with lumbar fusion surgery and receiving Washington State workers compensation followed by for a mean of 6.6 years.[60] This study reports that psychological comorbidities, characterized as including depression, dysthymia, manic-depressive disorders, stress, affective psychoses, or adjustment disorders, were associated with a higher risk of disability 2 years after lumbar fusion (OR 1.51 [1.05-2.26]).[60]

Key Question #5: What are the costs and potential cost-effectiveness of lumbar fusion relative to alternative treatment approaches?

Economic evaluations of lumbar spinal fusion in patients with uncomplicated DDD are limited both in number and in quality. Available evidence on the costs of lumbar fusion surgery suggest that in-hospital costs alone can approach $100,000 in the U.S., particularly for more complex forms of surgery. The results of two RCT-based economic evaluations mirrored findings for clinical outcomes. A comparison of fusion to interdisciplinary rehabilitation in which no material differences in clinical effectiveness were observed yielded a two-year cost-effectiveness estimate of >$100,000 per quality-adjusted life-year gained. A second comparison of fusion to variable approaches for physical therapy produced calculated cost per unit improvement in pain and function as well as per case of symptom improvement or return to work rather than traditional cost-effectiveness measures such as unadjusted or quality-adjusted survival. Finally, a survey-based study of low back pain patients' willingness to pay for surgery indicated a willingness to pay more than the actual observed costs of surgery for discectomy and decompression alone, but not for lumbar fusion.

While many studies in the available literature have documented increases in both the utilization and costs of lumbar fusion surgery, relatively few have focused specifically on costs and potential cost-effectiveness in the target population for this assessment—patients with degenerative disc disease and chronic low back pain not attributable to other conditions (e.g., severe stenosis, acute trauma, etc.) and without radiculopathy. We summarize the available economic evidence for patients with uncomplicated DDD below, as well as those from selected other studies commenting on cost data and/or trends

relevant to fusion surgery. Costs are presented in terms of 2014 US dollars, and were updated as necessary based on the medical care component of the U.S. Consumer Price Index.[73]

Utilization and Costs of Fusion in the U.S.

Given the policy interest around the use and appropriateness of fusion procedures in the U.S., it is not surprising that utilization of these procedures has been closely tracked. We chose to focus on comprehensive evaluations that have been performed most recently. One such study focused specifically on the use of lumbar fusion for DDD employed the Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample to evaluate trends from 2000-2009.[74] Population-adjusted utilization of fusion surgery increased 2.4-fold during this period, with the greatest increases seen in anterior approaches to fusion. Another relatively recent study used Medicare claims data to examine trends from 2002-2007 in utilization, outcome, and cost, although the focus of attention in this evaluation was on patients with spinal stenosis.[15] Results suggested a more than 15-fold increase (from 1.3 to 19.9 per 100,000 beneficiaries) in the rate of "complex" fusion procedures (i.e., more than two disk levels or a combined anterior/posterior approach), and an incidence of life-threatening complications with complex fusion (5.6%) more than 2-fold higher than among patients undergoing decompressive surgery without fusion. Adjusted hospital charges (in 2014 USD) ranged from $27,480 for decompression alone to $67,773 to complex fusion to $92,766 for complex fusion.

Martin and colleagues also explored whether differences in worker's compensation coverage policy for lumbar fusion in a variety of degenerative conditions had an impact on utilization and costs.[75] State inpatient databases were compared for California, which requires coverage in any situation in which a second opinion agrees with the first, and Washington, which applies utilization review criteria, requires imaging confirmation of spinal instability, and limits the initial procedure to a single disc level. In 2008-2009, the age- and sex-adjusted rate of lumbar fusion in the worker's compensation population was 19.0 per 100,000 employed adults in California and 12.9 per 100,000 in Washington ($p<0.001$). Rates of reoperation and readmission within three months of the initial procedure were also statistically-significantly higher in California. Finally, after adjustment for age, sex, comorbidity, and indication for fusion, mean hospitalization costs (2014 USD) were over 20% higher in California ($59,168 versus $48,271 for Washington, $p<0.001$).

Cost-Effectiveness of Lumbar Fusion in DDD

Two of the RCTs summarized in our assessment featured within-trial economic evaluations. In one, Rivero-Arias and colleagues evaluated the cost-effectiveness of lumbar fusion over a 2-year period[76] based on clinical, utility, and micro-costed data collected during Fairbank's RCT comparing lumbar fusion to intensive rehabilitation.[24] Costs were calculated based on itemized resources and unit costs for surgical, rehabilitation, and follow-up services utilized. Utility estimates were based on direct collection of data from the EuroQol EQ-5D questionnaire at multiple timepoints. Interestingly, while productivity loss was also costed, these estimates do not appear to have been used in the evaluation, which is described as having been conducted from the perspective of the British National Health Service.[76]

Two-year costs for surgery and rehabilitation (in 2014 US dollars) totaled $18,345 and $10,604 respectively. The difference in quality-adjusted survival between groups was 0.068 in favor of surgery (although this was not a statistically-significant difference). Cost-effectiveness (2014 USD) was $113,838 per quality-adjusted life year (QALY) gained for surgery. The authors concluded that such a ratio would not represent a cost-effective use of resources over a 2-year window, and sensitivity analyses suggested

that cost-effectiveness might only be approached if differences in utility persisted over the long term and/or greater than 20% of rehabilitation patients opted for surgery each year.

The other trial-based evaluation comes from the Swedish Lumbar Spine Study[25] and also involved costing of resources consumed during the 2-year study.[77] Unfortunately, cost-effectiveness was expressed not in terms of cost per QALY or life-year gained, but in terms of unit improvements in disability, treatment success, and return to work. In primary analyses, cost-effectiveness of fusion (in 2014 USD) versus usual care was estimated to be $2,363 per unit improvement on the ODI. Original cost-effectiveness calculations appeared to treat differences in return to work and significant clinical improvement as whole numbers rather than proportions. When considered as proportions (i.e., differences in the probability of these outcomes), cost-effectiveness was $54,527 per significant clinical improvement and $81,011 per return to work respectively.

We identified two additional cost-effectiveness evaluations that made use of clinical data, although not from studies that were considered for our evidence review. Adogwa and colleagues examined the cost-effectiveness of TLIF in 45 patients with grade 1 spondylolisthesis[78], while Glassman et al. assessed the cost-effectiveness of PLIF among patients with DDD as well as other conditions (e.g., disc herniation).[79] In both studies, however, costs and QALYs at two years were compared to those before surgery in the same population rather than to a control group receiving a contemporaneous intervention. In Adogwa's study, cost-effectiveness was estimated to be $46,428 per QALY gained (2014 USD) at two years.[78] In the Glassman evaluation, the cost-effectiveness of fusion (2014 USD) was $34,565 per QALY gained when only direct health care costs were considered and $56,443 per QALY gained when costs of lost productivity were added. Again, these ratios are calculated in relation to a pre-surgical state rather than to the costs and outcomes associated with an alternative treatment.

Other Economic Evaluations

Fayssoux and colleagues estimated the indirect costs associated with surgery for single-level DDD by using pooled data from an RCT of lumbar fusion and artificial disc replacement.[80] In the first year postoperatively, rates of full- or part-time employment declined from approximately 54% at baseline to less than 30% at 6 weeks, but returned to baseline levels by one year. Lost wages totaled approximately $2,900 per patient in the first year. By the end of the second year of follow-up, 63% of patients reported full- or part-time employment.

Another study involved the use of a post-surgery evaluation of the value that patients ascribe to individual surgical procedures for low back pain.[81] A total of 115 Swiss patients who had undergone discectomy, decompressive surgery, or fusion for a variety of degenerative lumbar conditions were surveyed regarding the maximum they would be willing to pay for each of these procedures, controlling for other factors such as satisfaction with the procedure, family income, and other financial resources. For both discectomy and decompression, the maximum willingness-to-pay (WTP) threshold for surgery was higher than the actual cost of the surgical procedures. For lumbar fusion, however, patients reported a maximum willingness-to-pay level of $19,712 (2014 USD), compared to an actual average hospital cost of $24,676 ($p<0.05$).

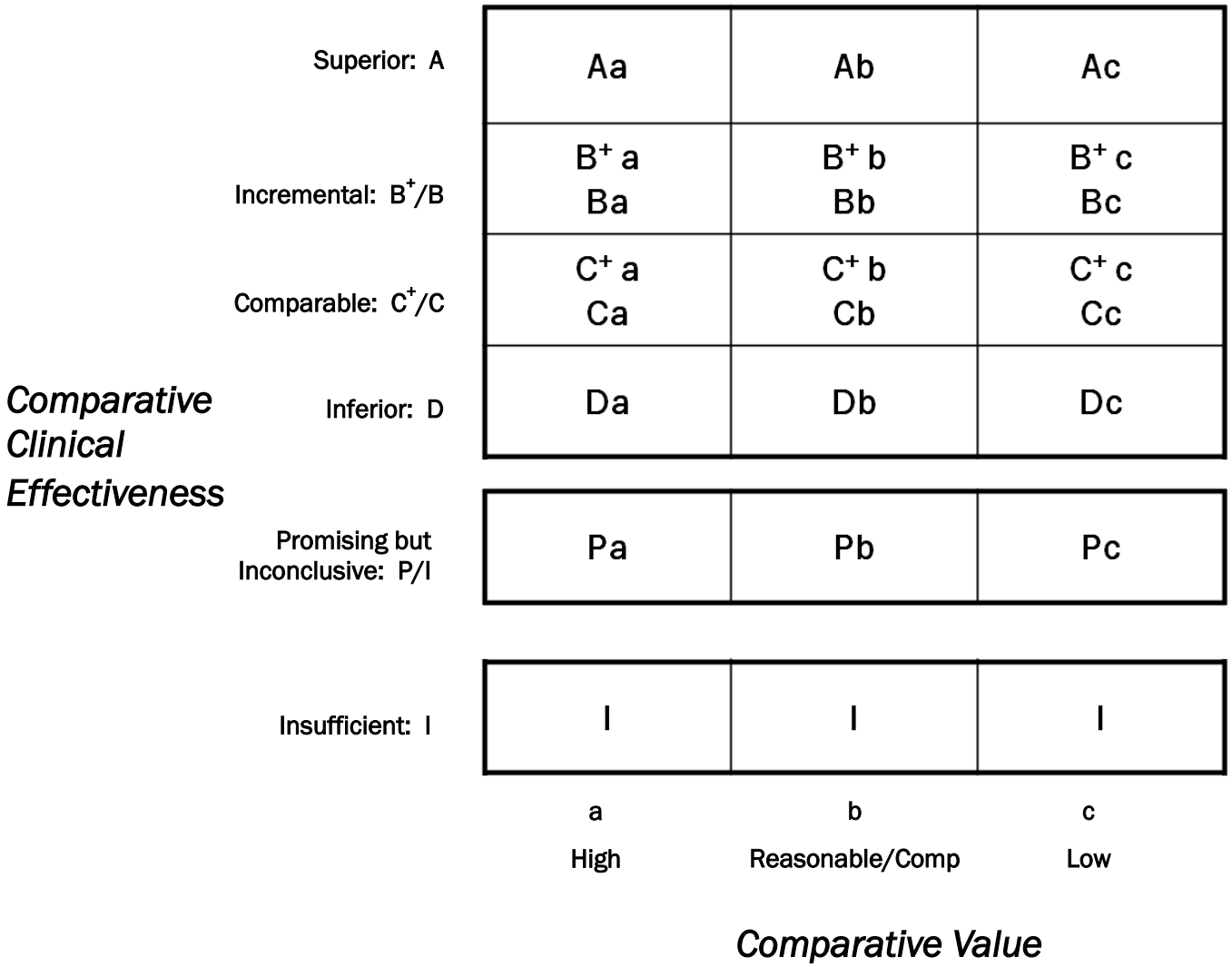
Finally, Alvin and colleagues conducted a systematic review to document variation in cost-calculation methods in economic evaluations of cervical and spinal lumbar surgery.[82] A total of 37 economic evaluations were identified. Sources of costs varied widely, with approximately one-third of evaluations using public-payer reimbursement, another one-third based on procedure micro-costing approaches, and the remainder using cost-to-charge ratios or other government data sources. Of perhaps greater

concern, one-quarter of the cost-effectiveness evaluations that stated use of a societal perspective did not include calculations of indirect costs, and there was great variation in the types of direct costs considered.

ICER Integrated Evidence Ratings

The ICER integrated evidence rating matrix is shown below; a detailed explanation of the methodology underpinning this rating system can be found in Appendix C to the full report. Separate ratings are provided for each of the populations and procedure comparisons under consideration; the ratings and rationale are described on the following pages.

Figure ES-3: ICER Integrated Evidence Ratings



Lumbar Spinal Fusion vs. Conservative Management in Uncomplicated DDD

1. Lumbar Fusion vs. Intensive/Interdisciplinary Rehabilitation: Dc ("Inferior/Low Value")
2. Lumbar Fusion vs. Less Intensive Conservative Management: Cc ("Comparable/Low Value")

Rationale for ICER Ratings

As noted in this review, there were clear distinctions in the available evidence comparing lumbar fusion to conservative management in patients with degenerative disc disease and no other acute, systemic, or clearly anatomic causes for low back pain (i.e., "uncomplicated DDD"). Randomized and other studies comparing fusion to structured, interdisciplinary rehabilitation programs that typically add educational and behavioral components to physical and exercise therapy show no clinical benefit for surgery in these comparisons. In addition, regardless of the comparator, lumbar fusion procedures are associated with relatively high rates of overall complications and reoperation. In our view, combining the evidence on clinical benefits and harms for fusion yields a net health benefit rating of "Inferior" in comparison to interdisciplinary rehabilitation. In addition, while there is a lack of high-quality evidence on cost-effectiveness in this setting, fusion appears to represent a high-cost intervention in the U.S. for no material gain in relation to interdisciplinary programs (i.e., of "low" value).

The evidence for lumbar fusion in comparison to less-intensive, often single-component conservative care (e.g., physical or exercise therapy alone) involves a more complex tradeoff. Randomized evidence suggests statistically-significant (if not consistently clinically-significant) improvements in pain, function, and return to work for versus conservative management in these settings, but with the notable caveat that short-term benefits (i.e., at 1-2 years) appear to diminish with longer-term follow-up. These findings are further complicated by high rates of crossover between treatment groups in some studies. Given fusion's potential for harm, we feel that the available evidence suggests that fusion and less-intensive conservative management are "functionally equivalent" (i.e., a rating of "Comparable" on the ICER matrix). Available economic evidence in these settings is limited and subject to quality concerns (e.g., use of nontraditional measures of cost-effectiveness, comparisons to pre-surgical states rather than contemporaneous control therapy). Given fusion's high cost and only modest long-term benefit for these comparisons, we consider fusion also to be of low value when compared to less-intensive conservative management.

1. Background

Condition

Low back pain is an exceedingly common complaint and a substantial cause of disability. At any given point in time, more than 10% of individuals are diagnosed with low back pain, and lifetime prevalence ranges from 60-70% in industrialized countries such as the US.[1] The economic impact of low back pain is also substantial. It is the second most common reason for all physician visits in the U.S.[2], and is responsible for approximately $30 billion in direct medical costs annually[3]. In addition, low back pain is associated with substantial indirect costs, in large part due to its detrimental impact on productivity; it is estimated that over 3% of the U.S. work force is compensated for back pain or injury each year[4], with approximately 187 million missed work days and wage losses accounting for an additional $22.4 billion in annual indirect costs[5].

With low back pain often presenting as a temporary condition, and an estimated 25-58% of cases spontaneously resolving[6], nonsurgical, i.e. conservative, treatment is the primary treatment modality at diagnosis. Conservative treatment may include any number of non-surgical therapies, in a structured or unstructured setting, and to lesser or greater degrees of intensity; such therapies include exercise, physical therapy, education, cognitive behavioral therapy, acupuncture, or spinal manipulation. However, persistent low back pain that is refractory to conservative treatment may be seen in as many as one-quarter of patients six months after an initial episode.[7]

Low back pain can be caused by a number of specific and nonspecific conditions, all of which differ in prevalence and affect different age groups. Nerve irritation, muscle strain, and bone or soft tissue damage may all give rise to low back pain. Another common cause of low back pain is lumbar degenerative disc disease (DDD), arising from natural degeneration of an intervertebral disc. DDD is commonly associated with low back pain in many individuals. Use of the term “disease” to describe this condition is something of a misnomer, however, as disc degeneration (dehydration and shrinkage) is a natural consequence of aging, and many individuals never develop overt symptoms of DDD; it is the symptoms of DDD (e.g., pain, limited mobility) that are the primary causes of concern. Diagnosis and subsequent treatment typically involves an initial history and physical examination by a clinician. Depending on the presentation, the clinician might prescribe various conservative self-care therapies or will perform a diagnostic exam to check the patient’s pain tolerance, functional capabilities, and reflexes.[8] An MRI and/or CT scan may be used to identify other potential anatomic causes of the patient’s symptoms, including other co-occurring conditions such as radiculopathy (compression of the root nerve), spondylolisthesis (slippage of a vertebral disc over another, causing spinal instability), or spinal stenosis (narrowing of the spinal canal), lumbar disc herniation (the rupture of an intervertebral disc which then pushes outside its normal boundary).[9,10] The process of disc degeneration appears to be influenced by demographic and behavioral factors (e.g., age, occupation, and activity level), lifestyle (e.g., obesity, smoking), and importantly, genetics.[6]

Multiple treatment options are available for symptoms associated with DDD of the lower back, including conservative measures, minimally-invasive treatments such as spinal injections and radiofrequency ablation, and surgical intervention. Conservative, non-invasive approaches vary widely in method and intensity, and are typically used as a first-line treatment approach for patients complaining of low back pain. When pain becomes chronic (i.e., continues for longer than three months), more intensive

conservative management using interdisciplinary methods is often considered. If these are unsuccessful, management with surgery can be considered. Lumbar fusion surgery, which involves the creation of a permanent connection across the vertebral space by means of a graft, is often considered when conservative treatments fail to relieve the patient's pain.[11] However, many patients may be at risk of continued persistent low back pain, as initial surgery is subject to high rates of reoperation with declining success rates after each consecutive surgery. It is estimated that as many as 80,000 cases of so-called "failed back surgery syndrome" are seen in the U.S. each year, although this figure includes not only fusion but other forms of back surgery.[12]

Not surprisingly, there is significant interest on the part of patients, clinicians, policymakers, and other stakeholders in evaluating the clinical and economic impact of lumbar fusion for patients with chronic low back pain and DDD. An evidence-based inquiry into lumbar fusion as a treatment option for DDD is complicated by the fact that there exists no consensus regarding a true "gold standard" treatment for DDD. Given that lumbar fusion is commonly employed intervention for a number of indications (representing 3.1% of all surgical procedures in the US), a careful evaluation of its effectiveness relative to conservative treatment of DDD will serve to inform policy around its use.

While some treatment options are used exclusively in certain patient populations, they can be generally characterized as follows:

- Non-surgical
    - Conservative treatment
        - Lower intensity: simple, unimodal conservative treatment. This includes medications, physical and/or exercise therapy, behavioral therapy, chiropractic, alternative therapy (e.g., acupuncture, yoga)
        - Higher intensity: Interdisciplinary rehabilitation. This includes intensive, multimodal rehabilitation that is physician-directed and may include workplace, exercise, educational, and/or psychologist- or therapist-led behavioral interventions
    - Minimally invasive procedures
        - Spinal injections (e.g., epidural steroids, facet joint)
        - Radiofrequency denervation
        - Intradiscal electrothermal therapy: Passage of a catheter into the lumbar disc space, and heating up the outer core.
    - Posterior dynamic stabilization: Devices are used to preserve motion in the spine while also releasing pressure on the degenerated disc. The devices can be used in conjunction with fusion, or as stand-alone treatments.
- Surgical
    - Discectomy[10]
    - Spinal fusion: the intensity of each of the below procedures is associated with the number of vertebral segments involved, with the fusion of 2 segments limiting motion more so than 1.
        - Posterolateral gutter fusion - the procedure is done through the back, and involves placement of the bone graft between the transverse processes, commonly using pedicle screw fixation.

- Interbody fusion: the procedures are done using various approaches, and involve removing the disc between two vertebrae and inserting the bone graft into the space created between the vertebral bodies
    - Posterior lumbar interbody fusion (PLIF) - the procedure is done from the back
    - Transforaminal lumbar interbody fusion (TLIF) - this procedure is done from the back of the spine, differing from the PLIF mainly in the angle at which the disc is approached
    - Anterior lumbar interbody fusion (ALIF) - the procedure is done from the front
    - Extreme Lateral Interbody Fusion (XLIF) - an interbody fusion in which the approach is from the side

Policy Context

Due to the prevalence of low back pain and the varying nature of the conditions that underlie it, numerous management options are available. These options vary substantially in intensity of treatment and follow-up, degree of invasiveness, and most importantly, level of evidence regarding their effectiveness. Although there is lack of consensus on when lumbar fusion surgery is indicated, how the surgery should be performed, and long-term prognosis after surgery[83], the number of lumbar fusion surgeries performed in the U.S. has nevertheless increased more than two-fold between 2000 and 2009[74]. In particular, some studies have shown poor success rates for lumbar fusion when used to treat low back pain associated specifically with uncomplicated disc degeneration.[84] Not surprisingly, there is significant interest on the part of patients, clinicians, policymakers, and other stakeholders in evaluating the clinical and economic impact of lumbar fusion for patients with chronic low back pain and DDD.
An evidence-based inquiry into lumbar fusion as a treatment option for DDD is complicated by the fact that there exists no consensus regarding a true “gold standard” treatment for DDD. Building such consensus around treatment requires comparison of surgical to conservative treatment modalities, providing a clear picture of all treatment options, and allowing for evidence-based evaluation. Lumbar fusion is a commonly employed method of surgical intervention for DDD, and a careful evaluation of its effectiveness relative to conservative treatment modalities will serve to inform policy around its use.

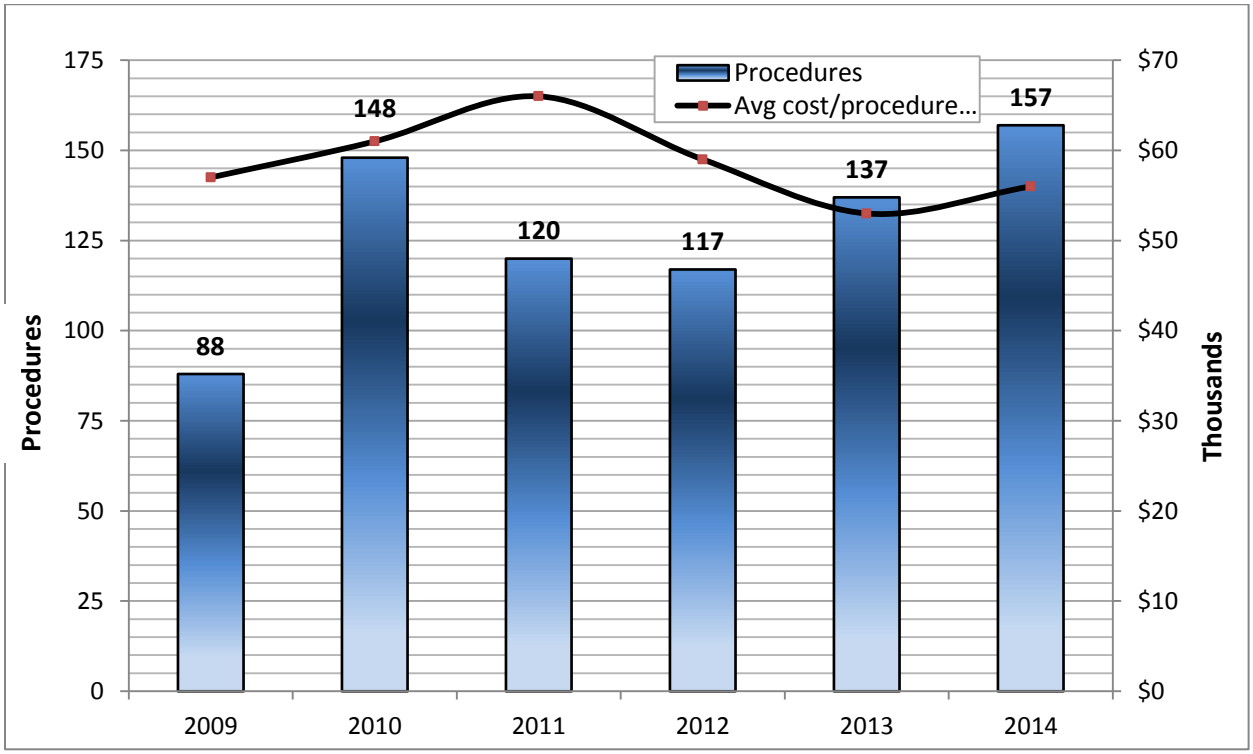
2. Washington State Agency Utilization Data

LUMBAR FUSION The lumbar fusion study includes members from both the Public Employee Benefits (PEBB) and members of Medicaid. The study periods for the populations are: PEBB calendar year 2009 through 2014; Medicaid calendar year 2012 – 2014. The primary study inclusion criteria included a CPT Code of 22533, 22558, 22612, 22630, 22633, or 22849. Cost included all professional, inpatient, and ancillary claims for the CPT as the first date of service. Finally, claims that included a $0 allowed amount and a $0 paid were excluded as denied claims.

Public Employee Benefits (PEBB)
Utilization: Lumbar Fusion
2009 – 20014
NOTE: *Submitted, Allowed and Paid Dollars are rounded*

Year	Unique Patients	Submitted Amt	Allw Amt	PEBB/UMP Paid$	Average PEBB/UMP Paid$/Patient
2009	85	$9,952,000	$5,011,000	$4,912,000	$57,788
2010	145	$19,126,000	$8,960,000	$8,793,000	$60,641
2011	119	$20,528,000	$7,951,000	$7,790,000	$65,462
2012	116	$15,420,000	$6,932,000	$6,832,000	$58,897
2013	136	$22,368,000	$7,223,000	$7,094,000	$52,162
2014	154	$26,612,000	$8,810,000	$8,680,000	$56,364

Public Employee Benefits (PEBB)
Utilization: Lumbar Fusion Annual Procedures and Average Allowed /Procedure Over Time
2009 – 20014



Public Employees Benefits (PEBB)

Utilization: Lumbar Fusion
Most Frequently Used ICD-9 Diagnoses Codes Utilized on Lumbar Fusion Claims

Diag Long Desc - Prin	2009	2010	2011	2012	2013	
ACQUIRED SPONDYLOLISTHESIS	28	59	53	65	68	81
SPINAL STENOSIS, LUMBAR REGION, WITHOUT NEUROGENIC CLAUDICATION	27	41	37	47	57	42
DEGENERATION OF LUMBAR OR LUMBOSACRAL INTERVERTEBRAL DISC	25	32	29	23	29	20
DISPLACEMENT OF LUMBAR INTERVERTEBRAL DISC WITHOUT MYELOPATHY	8	15	20	17	20	28
LUMBOSACRAL SPONDYLOSIS WITHOUT MYELOPATHY	9	18	`27	17	18	19

Diag Long Desc - Prin	2009	2010	2011	2012	2013	
SPINAL STENOSIS, LUMBAR REGION, WITH NEUROGENIC CLAUDICATION		2	8	17	23	28
SCOLIOSIS (AND KYPHOSCOLIOSIS), IDIOPATHIC	3	4	6	11	7	10
MECH COM ORTH DEV NEC	3	5	9	8	3	10
POSTLAMINECTOMY SYNDROME, LUMBAR REGION	3	2	3	6	4	7
OTHER KYPHOSCOLIOSIS AND SCOLIOSIS	2	3	3	3	4	2

Utilization for the Medicaid population included: two members with Lumbar fusions in each of the three years examined; and 20 members with two lumbar surgeries over the three years. A handful of members had more than one surgery during the same year. The balance of the population had a single surgery.

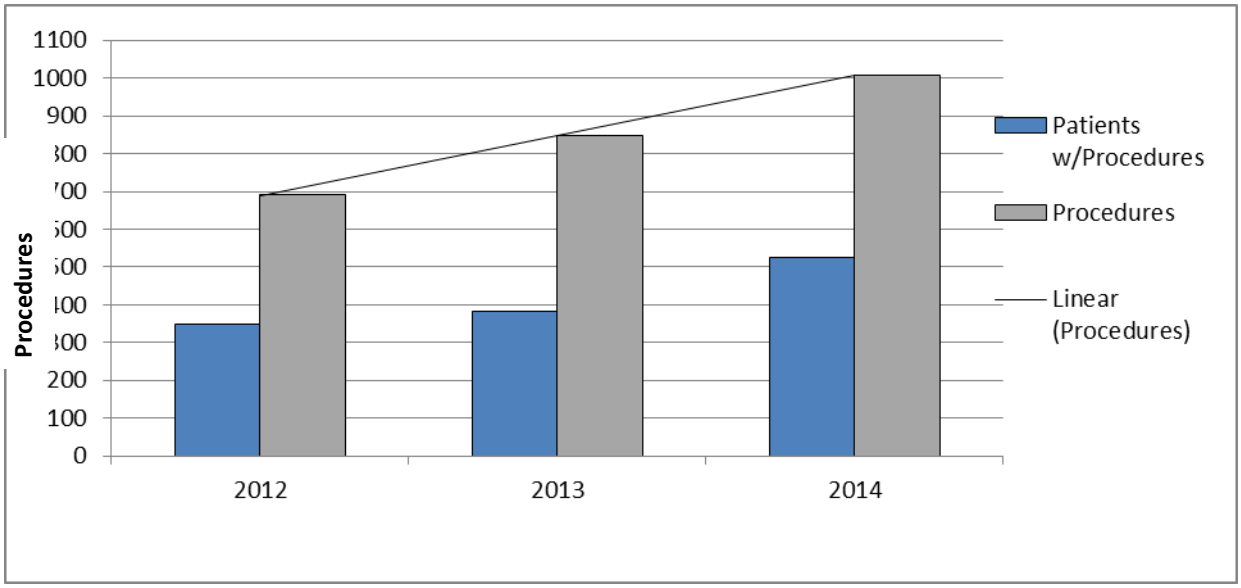
Medicaid

Utilization: Lumbar Fusion

2012 – 20014

Year	Patients w/Procedures	Unique Patients *(Non-repeating in the period)*	Procedures
2012	347	328	690
2013	383	325	848
2014	524	502	1009

Medicaid
Utilization: Lumbar Fusion with Linear Trend Line
Annual Number of Procedures, Patients w/procedure (not mutually exclusive)
2012 – 20014



Medicaid
Utilization: Lumbar Fusion
Most Frequently Used ICD-9 Diagnoses Codes on Lumbar Fusion Claims

		Count of Members		
Header Dx	Header Dx Code	2012	2013	2014
7384	Acq spondylolisthesis	134	128	257
72402	Spin sten,lumbr wo claud	111	184	206
72252	Lumb/lumbosac disc degen	42	39	69
7213	Lumbosacral spondylosis	28	53	54
72210	Lumbar disc displacement	25	29	51
72403	Spin sten,lumbr w claud	17	33	32
71888	Jt derangment NEC-oth jt	19	24	29

		Count of Members		
Header Dx	Header Dx Code	2012	2013	2014
7244	Lumbosacral neuritis NOS	23	17	29
75612	Spondylolisthesis	7	21	37
72283	Postlaminect synd-lumbar	15	26	22

Procedure Codes used for Lumbar Fusion Analysis

Proc Code	Proc Long Desc
22533	LAT EXTRACAVITARY ARTHRODESIS, W/MIN DISKECTOMY; LUMBAR
22558	ARTHRODESIS W/MINI DISKECT; LUMB
22612	ARTHRODESIS-POST/POSTLAT-1; LUMB
22630	ARTHRODESIS-POST INTERBODY-1; LUMB
22633	ARTHDSIS POST/POSTEROLATRL/POSTINTERBODY LUMBAR
22849	REINSERTION SPINAL FIXA DEVICE

The major approaches to lumbar spinal fusion and conservative, nonsurgical management are described in further detail below. Of note, other minimally-invasive procedures (e.g., spinal injections, denervation procedures) are used in patients with uncomplicated DDD but are not described here given the contrast of primary interest for our review.

Lumbar Spinal Fusion

During spinal fusion procedures, the spine is stabilized by fusing two or more vertebrae together, using metal rods, bone grafts, or screws.[85] Spinal fusions are classified as either simple (1 or 2 disc levels or a single surgical approach) or complex (more than 2 disc levels or a combined anterior and posterior approach). Fusion may or may not use instrumentation such as screws, plates, or cages. Instrumentation is generally used as an internal splint to hold the vertebrae together while the bone grafts heal. Bone or bone substitutes are used to help fuse the vertebrae together. The bone may be taken from another bone in the patient (autograft) or from a bone bank (allograft). Bone morphogenic proteins may also be used as an alternative to autograft.

During lumbar fusion, the surgeon removes the lamina to help relieve the pressure on the nerve. The surgeon then removes any additional bone that may impinge upon the affected nerve. Bone grafts are

then added to the spine; these will eventually fuse with the spine to form a solid bone. Instrumentation may be added to provide additional stability while the grafts heal. There is generally more discomfort experienced after fusion surgery compared to other procedures and recovery takes much longer. Patients usually stay in the hospital for at least three to four days post-procedure. Substantial bone healing takes some time to achieve and the healing process varies from person to person. The indication of bone healing, as evidenced by an X-ray, is not attempted until approximately 6 weeks post-procedure. During this time, the patient's activity must be limited. The surgeon may recommend a post-operative rehabilitation program.[14]

Risks associated with spinal fusion include nerve root damage, bleeding, and infection. While the major risks are relatively rare, the odds of injury are higher with increasing complexity of surgical approach and use of instrumentation.[15] Other complications, common to all types of major surgery, may include blood clots, myocardial infarction, pulmonary embolism, and pneumonia.

The main approaches to lumbar fusion surgery are as follows:

Posterolateral fusion (PLF)

In a posterolateral fusion, the surgical approach to the spine is from the back through a midline incision that is approximately three inches to six inches long. A bone graft is obtained and laid out in the posterolateral portion of the spine. This region lies on the outside of the spine and is a very vascular area, which is important because the fusion needs blood to supply the nutrients for it to grow. A small extension of the vertebral body in this area (transverse process) is a bone that serves as a muscle attachment site. The large back muscles that attach to the transverse processes are elevated up to create a bed to lay the bone graft on. The back muscles are then laid back over the bone graft, creating tension to hold the bone graft in place. This approach is often considered the "gold standard" for spinal fusion surgery.

Interbody Fusions

Designed to be a less invasive way of obtaining a spinal fusion by using two threaded titanium cylinders to hold the vertebrae in proper position while the spine fusion occurs. These procedures are done using various approaches, and involve removing the disc between two vertebrae and inserting the bone graft into the space created between the vertebral bodies. They are described in detail below:

Posterior lumbar interbody fusion (PLIF)

Unlike the posterolateral fusion, the PLIF achieves spinal fusion in the low back by inserting a cage made of either allograft bone or synthetic material (PEEK or titanium) directly into the disc space. PLIF surgery has a higher potential for a solid fusion rates than posterolateral fusion rates because the bone is inserted into the anterior portion (front) of the spine.

Anterior lumbar interbody fusion (ALIF)

The anterior lumbar interbody fusion is similar to the PLIF approach, except that in the ALIF, the disc space is fused by approaching the spine through the abdomen instead of through the lower back. A three-inch to five-inch incision is made on the left side of the abdomen and the abdominal muscles are retracted to the side.

Transforaminal lumbar interbody fusion (TLIF)

TLIF fuses the anterior (front) and posterior (back) columns of the spine through a single posterior approach. This procedure is done from the back of the spine, differing from the PLIF mainly in the angle at which the disc is approached.

Extreme Lateral Interbody Fusion (XLIF)

An interbody fusion approach in which the surgeon accesses the intervertebral disc space and fuses the lumbar spine using a surgical approach from the side (lateral) rather than from the front (anterior) or the back (posterior).

Conservative, Nonsurgical Management

Conservative, non-invasive approaches vary widely in method and intensity. Further detail on this variability is available in the evidence review. Lower intensity treatments typically include medications, physical and/or exercise therapy, behavioral therapy, chiropractic, and alternative therapy (e.g., acupuncture, yoga). These are typically used as a first-line treatment approach for patients complaining of low back pain. When pain becomes chronic (i.e., continues for longer than three months), interdisciplinary rehabilitation is often considered. Interdisciplinary rehabilitation programs are interventions that combine and coordinate physical, vocational, and behavioral components.[16] These programs are typically physician-directed, with care provided by multiple health care professionals with different clinical backgrounds. The intensity and content of interdisciplinary therapy varies widely; duration of treatment may be as short as one week or as long as 15 weeks and activity levels range from one to eight hours on any given day. Programs typically involve some component of group therapy, usually held in groups of up to 10. Interdisciplinary programs vary not only in duration and intensity, but also in the types of components provided. Worksite interventions, strength training, aerobic exercises, educational interventions, and psychological interventions are all examples of components that can constitute an interdisciplinary program.

4. Clinical Guidelines and Training Standards

American Association of Neurological Surgeons (AANS) (2014)

http://thejns.org/doi/pdf/10.3171/2014.4.SPINE14270
Lumbar fusion is recommended for patients with 1- or 2- level degenerative disc disease without stenosis or spondylolisthesis if chronic low back pain persists after conservative treatment, which may include physical therapy and other non-operative measures.

American Pain Society (APS) (2009)

http://journals.lww.com/spinejournal/Abstract/2009/05010/Interventional_Therapies,_Surgery,_and.14.aspx
For patients with non-radicular low back pain who have not responded to usual care, APS advises clinicians to consider intensive interdisciplinary rehabilitation that combines physical rehabilitation with a psychological and social or occupational component.

For patients with non-radicular low back pain, common degenerative spinal changes, and persistent and disabling symptoms, APS recommends that clinicians use a shared-decision making approach in deciding whether or not to pursue fusion surgery. Physicians should discuss with patients the similar efficacy of interdisciplinary rehabilitation, and the small to moderate average benefit of surgery over interdisciplinary rehab. If the patient and clinician together decide that surgery is the best option, instrumented fusion is associated with enhanced fusion rates over non-instrumented fusion, though the evidence is not sufficient to suggest better outcomes. No specific fusion method is recommended over another.

For patients with persistent non-radicular low back pain, APS found evidence to be insufficient to evaluate long-term benefits and harms of vertebral disc replacement, local injections, botulinum toxin injection, epidural steroid injection, intradiscal electrothermal therapy, therapeutic medical branch block, radiofrequency denervation, sacroiliac joint steroid injection, or intrathecal therapy with opioids or other medications. Facet joint corticosteroid injection, prolotherapy, and intradiscal corticosteroid injection are not recommended.

International Society for the Advancement of Spine Surgery (ISASS) (2015)

https://www.isass.org/public_policy/2011-07-15_policy_statement_lumbar_surgery.html
Degenerative disc disease is considered to be a medically necessary indication for fusion—at a maximum of two levels—when the following criteria are met:

- The patient is experiencing clinically significant pain and disability consistent with discogenic pain;
- Imaging studies suggest morphological disc degeneration;
- The patient has tried 6 consecutive months of structured conservative management, including pain medication, activity modification, and daily exercise, with demonstrated compliance, and has not shown sufficient improvement;
- Following 6 months of conservative management, the patient has tried intensive multidisciplinary rehabilitation if available and covered by the patient's insurance. The program

must include a cognitive/behavioral component, with at least 80 hours of on-site treatment during a 2-4 week period;

- The patient has been screened for possible mental illness or substance abuse issues and has undergone professional treatment if a condition is identified;
- The patient is not currently involved in an ongoing litigation case related to his or her back;
- The patient is between the ages of 25 and 65;
- The patient is not pregnant; and
- Provocative discography or magnetic resonance spectroscopy has been used to confirm that pain is likely due to disc degeneration observed on imaging.

National Institute for Health and Care Excellence (NICE) (2009)

http://www.nice.org.uk/guidance/cg88/chapter/1-guidance
For first line therapy, NICE advises clinicians to promote self-management and provide patients with strategies to manage their low back pain. Patients may also be offered medication, including NSAIDs, opioids, or antidepressants, as well as one of the following treatment options, depending on patient preference: a structured exercise program, a course of manual therapy, or a course of acupuncture, each lasting 12 weeks. If these therapies do not provide sufficient improvement, physicians may consider a combined physical and psychological treatment program that includes at least 100 hours of treatment over an 8 week period. If the patient has completed these steps and continues to have pain, referral to a specialist for spinal surgery may be considered.

Prior to surgery, any patient with psychological distress should receive treatment. Patients should be referred to a specialist, and physicians should consider all possible risks for the patient. Patients should not be referred for other procedures, including intradiscal electrothermal therapy, percutaneous intradiscal radiofrequency thermocoagulation, or radiofrequency fact joint denervation.

These guidelines were current as of 2009, and an update is currently in development for 2016.

North American Spine Society (2014)

https://www.spine.org/Documents/PolicyPractice/CoverageRecommendations/LumbarFusion.pdf
Lumbar fusion is indicated for discogenic low back pain secondary to a degenerated disc when the following criteria are met:

- Single level disease confirmed by MRI with moderate to severe degeneration of the disc with Modic changes;
- Patient has had symptoms for at least one year that have not responded to nonsurgical options, which at minimum must include physical therapy. Other nonsurgical options may include pain management, injections, cognitive behavioral therapy, and exercise programs;
- Patient does not have an active psychological disorder that requires pharmacologic management;
- Patient has not smoked for at least three months prior to surgery; and
- The primary complaint is axial pain, with a possible secondary complaint of pain in lower extremities.

5. Medicare and Representative Private Insurer Coverage Policies

5.1 Centers for Medicare and Medicaid Services (CMS)

There are currently no national or local coverage determinations for lumbar fusion that pertain to Washington State.

5.2 Representative National Private Insurer Policies

Aetna

http://www.aetna.com/cpb/medical/data/700_799/0743.html
Lumbar spinal fusion for degenerative disc disease is not covered due to a lack of evidence on effectiveness.

Anthem

https://www.anthem.com/ca/medicalpolicies/guidelines/gl_pw_c160722.htm
Lumbar fusion is not considered medically necessary for low back pain due to degenerative disc disease or degenerative lumbar spondylosis without stenosis or spondylolisthesis.

CIGNA

https://cignaforhcp.cigna.com/public/content/pdf/coveragePolicies
Single level lumbar fusion is covered for degenerative disc disease without instability when there is unremitting pain and functional impairment for at least 12 months and all of the following conditions are met:

- Continued pain and impairment despite at least 6 consecutive months of structured, physician supervised conservative management including exercise, medication, physical therapy, participation in at least 3 cognitive behavioral therapy sessions that address disease education, activity and lifestyle modification, and stress management
- Single level degenerative disc disease confirmed by imaging studies
- Statement from a licensed behavioral or medical health care provider attesting to an absence of underlying mental health issues that may contribute to chronic pain
- The individual does not smoke, or will refrain from smoking for at least 6 weeks prior to surgery

Humana

http://apps.humana.com/tad/tad_new/Search.aspx?criteria=spinal+fusion+surgery&searchtype=freetext&policyType=both
Lumbar fusion surgery is covered for a variety of conditions when confirmed by imaging studies. Covered indications include iatrogenic instability, severe degenerative scoliosis, spinal abscess or infection, spinal dislocation, spinal fracture, spinal stenosis associated with spondylolisthesis, spinal tuberculosis, spinal tumor, spondylolysis such as isthmic spondylolisthesis, and symptomatic pseudoarthrosis from a prior procedure. Many covered indications are subject to additional clinical criteria.

UnitedHealthcare

https://www.unitedhealthcareonline.com/ccmcontent/ProviderII/UHC/en-US/Assets/ProviderStaticFiles/ProviderStaticFilesPdf/Tools%20and%20Resources/Policies%20and%20Protocols/Medical%20Policies/Medical%20Policies/surgical_treatment_for_spine_pain.pdf
UnitedHealthcare does not have a plan-specific policy for lumbar fusion; enrollees must defer to their specific benefit document. The most recent update to their medical policy on surgical treatment for spine pain allows the use of extreme lateral interbody fusion (XLIF®) or direct lateral interbody fusion (DLIF), though no specific indications for these surgeries are listed.

5.3 Representative Regional Private Insurer Policies

Health Net

https://www.healthnet.com/static/general/unprotected/pdfs/national/policies/LumbarSpinalFusion.pdf
Lumbar fusion is also considered medically necessary for patients with chronic mechanical low back pain without the presence of radiographic intervertebral instability, when the patient has chronic, severe, and disabling pain from degenerative disc disease confirmed with CT or MRI. To be considered for fusion, that patient's pain must persist despite at least six consecutive months of non-surgical measures. Non-surgical measures may include reconditioning exercises, activity modification, physical therapy, or medications. The patient must be free from untreated underlying psychosocial issues and be motivated. All other possible sources of pain must be ruled out, and requests for fusion cannot be at more than 2 adjacent levels.

Premera Blue Cross

https://www.premera.com/medicalpolicies/CMI_125925.pdf
Premera Blue Cross considers lumbar fusion investigational for disc herniation, chronic nonspecific low back pain without radiculopathy, degenerative disc disease, initial discectomy/laminectomy for neural structure decompression, or facet syndrome. Patients must have participated in a physician supervised weight loss program lasting at least six consecutive months within the two years preceding surgery. Patients must also complete a psychological evaluation with a licensed mental health provider to assess emotional stability and ability to comply with post-surgical limitations.

The Regence Group

http://blue.regence.com/trgmedpol/surgery/sur187.pdf
Lumbar fusion is not considered to be medically necessary for disc herniation, degenerative disc disease with no radicular symptoms, initial discectomy/laminectomy for neural structure decompression, facet joint arthritis as a singular problem, or low back pain that does not meet other criteria.

6. Previous Health Technology Assessments and Systematic Reviews

We were able to identify four formal health technology assessments evaluating lumbar fusion surgery relative to conventional treatment, none of which found sufficient evidence for these comparisons. Many systematic reviews have evaluated RCT-based data for these interventions; only two recent systematic reviews have included data from all five published RCTs and are described in detail in the section below.

6.1 Health Technology Assessments

Agency for Healthcare Research and Quality (AHRQ, 2006):

http://www.cms.gov/determinationprocess/downloads/id41ta.pdf
The amount of evidence on lumbar spinal fusion does not demonstrate either short- or long-term benefits when compared with non-surgical treatment, especially for patients over 65 years of age, or for those with degenerative disc disease.

Agency for Healthcare Research and Quality (AHRQ, 2012):

http://effectivehealthcare.ahrq.gov/ehc/index.cfm/search-for-guides-reviews-and-reports/?pageAction=displayTopic&topicID=410

http://ebm.avalere.com/studies/24754?keywords=lumbar+fusion&saved_search_name=&utf8=%E2%9C%93
Limited evidence suggests that spinal fusion compared with physical therapy improves pain and function for adults undergoing fusion for low back pain due to disc degeneration. The incidence of adverse events (serious and minor) associated with fusion could also not be determined conclusively because of insufficient reporting and variation in surgical methods used in the different studies. The authors also noted that many of the studies reviewed were ultimately excluded for lack of relevance to modern treatment.

The Cochrane Collaboration (2005)

http://journals.lww.com/spinejournal/Abstract/2005/10150/Surgery_for_Degenerative_Lumbar_Spondylosis_.13.aspx
An updated Cochrane review found insufficient evidence on the effectiveness of anterior, posterior, or circumferential fusion for degenerative lumbar spondylosis and any fusion procedure relative to conventional physiotherapy or an exercise and rehabilitation program.

Washington State Health Care Authority (HCA, 2007)

http://www.hca.wa.gov/hta/Documents/lumbar_fusion_final_report_101907.pdf
A health technology assessment conducted by the ECRI Institute for the Washington State HCA found insufficient evidence on outcomes of lumbar fusion relative to conservative treatment, including intensive exercise/rehabilitation plus CBT or non-intensive physical therapy, in patients with or without prior back surgery.

6.2 Systematic Reviews

Phillips 2013[86]

http://journals.lww.com/spinejournal/Abstract/2013/04010/Lumbar_Spine_Fusion_for_Chronic_Low_Back_Pain_Due.18.aspx
Phillips and colleagues identified a total of six publications with 547 fusion and 372 conservative patients. The weighted average improvement on the ODI was 13.9 ± 8.7/100 (29% change; 95% CI: 18.7, 39.4) in the surgical group and 8.2 ± 6.2/100 (17.5% change; 95% CI: 8.5, 26.6) in the conservative group. The weighted average improvement in patient satisfaction was 74.8% (95% CI, 72.2, 77.4) in the surgical group and 55.6% (95% CI, 53.3, 57.9) in the conservative group, with an average reoperation rate for fusion of 7% (95% CI: 5.7, 8.3). The authors concluded that the literature supports fusion as a viable option for patients with a diagnosis of disc degeneration who are refractory to conservative care. However, this review has been criticized for not reporting the methodological approach used to conduct the meta-analysis, and for using duplicated study samples in their assessment of fusion relative to non-operative care[43].

Bydon 2014[41]

http://journals.lww.com/jspinaldisorders/Abstract/2014/07000/Lumbar_Fusion_Versus_Nonoperative_Management_for.9.aspx
A systematic review and meta-analysis evaluated 5 RCTs comparing lumbar fusion to conservative care. Bydon et al. observed that despite statistically-significant improvements in favor of surgery in three of these studies, the pooled data did not reveal a statistically-significant difference compared to the non-operative group (TE: -7.39; 95% CI: -20.26, 5.47). The authors were also unclear if the treatment effect in favor of surgery would lead to a clinically-significant difference. Notably, this review only considered changes on the ODI; data on other pain measurements (e.g., VAS) and patient satisfaction were not pooled.

7. Ongoing Clinical Trials

We did not identify any ongoing RCTs or observational studies comparing lumbar fusion to conventional treatment or minimally-invasive approaches for patients with uncomplicated DDD. The majority were long-term safety and tolerability studies of various instrumentation devices (frequently sponsored by the manufacturers) used in fusion surgery, comparisons to other surgical interventions (e.g., discectomy, total disc replacement), or the use of various bone graft material to improve fusion rate. The ongoing trials below represent a snapshot of those studies that most closely resemble the patient population of interest to this review.

Table 1: Ongoing Clinical Trials

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
Study to Evaluate Safety and Effectiveness of Dynamic Stabilization Versus Lumbar Fusion in Treatment of Multilevel Lumbar Disc Degeneration Disease (MLIDH) NCT02385695	Case-control	Posterior Dynamic Stabilization Internal Fixation and Fusion	N=102 Age 30-75 Participants will have multi-level lumbar disc degeneration disease and be scheduled for 2- or 3-level lumbar discectomy from L1 to S1 with or without dynamic stabilization or fusion. ODI scores should be at least 30% prior to surgery, and clinical symptoms must be consistent with a diagnosis of lumbar DDD.	Range of motion in lumbar spine at 24 months	August 2021
Posterior Dynamic Stabilization Versus Fusion in the Treatment of Lumbar Degenerative Disease (DYNORFUSE)	RCT	Posterior Dynamic Stabilization Fusion	N=440 Age >18 Participants must have a mono- or bi-segmental symptomatic lumbar degenerative disease with or without stenosis; an indication for fusion with spondylolisthesis	Difference in Oswestry disability index (ODI) between treatment groups at 2 years post intervention	November 2015

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
NCT01365754			of at least 5mm or segmental vertebral motion of at least 3mm or 10º on flexion/extension radiographs, or ii) predominant low back pain in combination with Modic changes; and failure of adequate conservative measures for more than 3 months.		
A Multi-Center Prospective Randomized Study Comparing Supplemental Posterior Instrumentation, Aspen™ Spinous Process System Versus Pedicle Screw Fixation, in Lateral Lumbar Interbody Fusion (LLIF) or Anterior Lumbar Interbody Fusion (ALIF) NCT01549366	RCT	Fusion with Aspen™ device Posterior fusion with pedicle screw instrumentation	Age 18-75 Up to 25 sites Diagnosis of primary symptomatic Degenerative Disc Disease (DDD) and/or spondylolisthesis confirmed with appropriate imaging studies and/or positive lumbar discography and an Oswestry Disability Index (ODI) v2.1 score >30%, and failed at least 3 months of conservative care (non-surgical) OR has clinical signs of neurological deterioration	Absolute change in Oswestry Disability Index (ODI) at 2 years post-operative	December 2015

8. Methods

Objectives

The primary objectives of the systematic review were to answer the following key questions, using the listed sources of evidence:

1. What is the comparative clinical effectiveness of lumbar fusion surgery for patients with chronic low back pain and uncomplicated DDD relative to that of conservative management, minimally-invasive treatments, and other nonsurgical approaches?
   - Sources: RCTs, high-quality comparative cohort studies, and high-quality systematic reviews of lumbar fusion vs. the comparators of primary interest

2. What are the rates of "treatment success" or "successful clinical outcome" of lumbar fusion as defined by measures of clinically-meaningful improvement in pain, function, quality of life, patient satisfaction, and/or work status?
   - Sources: RCTs, high-quality comparative cohort studies, and high-quality systematic reviews of lumbar fusion vs. the comparators of primary interest

3. What are the rates of adverse events and other potential harms (perioperative, long-term adverse events, and reoperations) associated with lumbar fusion surgery compared to alternative treatment approaches?
   - Sources: RCTs, high-quality comparative cohort studies, and high-quality systematic reviews of lumbar fusion vs. the comparators of primary interest; selected case series

4. What is the differential effectiveness and safety of lumbar fusion according to factors such as age, sex, race or ethnicity, pre-existing conditions (e.g., smoking history), intensity of conservative management (e.g., interdisciplinary rehabilitation vs. physical and/or behavioral therapy alone) technical approach to fusion (e.g., posterolateral vs. interbody, minimally-invasive vs. open procedures), initial vs. repeat surgery, insurance status (e.g., worker's compensation vs. other), and treatment setting (e.g., inpatient vs. ambulatory surgery)?
   - Sources: RCTs, high-quality comparative cohort studies, and high-quality systematic reviews of lumbar fusion vs. the comparators of primary interest; selected non-comparative case series

5. What are the costs and potential cost-effectiveness of lumbar fusion relative to alternative treatment approaches?
   - Sources: Published economic evaluations, agency data

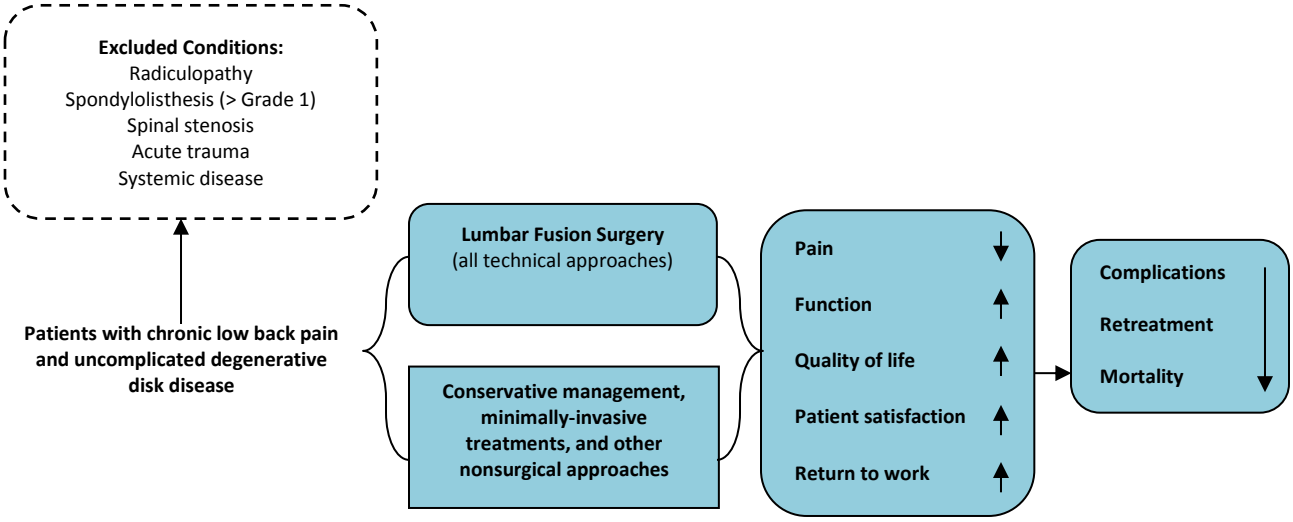
Analytic Framework

The analytic framework for this project is depicted on the following page. We expected that studies would vary substantially in terms of their entry criteria, as there is no agreed-upon standard of what constitutes uncomplicated lower back DDD. In addition, the fusion technique and intensity of the

conservative intervention may have differential effects on the outcomes of primary interest in low back pain studies, including pain, function, quality of life, patient satisfaction, and work status. Finally, RCTs of fundamentally different interventions (e.g., surgery for pain relief versus rehabilitation for functional restoration) may have difficulty enrolling and randomizing patients, resulting in many studies with inadequate statistical power or other quality concerns (e.g., high dropout and/or crossover rates).

There were expected limitations on the available evidence in terms of (a) comprehensive comparisons of lumbar fusion to conservative management, and (b) long-term data on effectiveness and potential harms. As such, judgments about the effectiveness of these interventions rested predominantly upon individual consideration of each type of surgery and its relevant comparators, evaluation of procedure-specific risks, and linkage of shorter-term outcomes to higher-quality data on long-term effects where available.

Figure 1. Analytical Framework: Lumbar Fusion



Population, Intervention, Comparators, and Outcomes, and Sources: PICOS

Specific details on the proposed scope (Population, Intervention, Comparators, and Outcomes, and Sources: PICOS) are detailed in the following sections.

Population

The target population for this review included adults (age >17 years) with chronic (≥3 months) low back pain and uncomplicated degenerative disc disease. Specifically, as in the original review, studies of patients with conditions such as radiculopathy, spondylolisthesis (> Grade 1) or severe spinal stenosis, as well as those with acute trauma or systemic disease affecting the lumbar spine (e.g., malignancy) were excluded. We recognize that some studies of lumbar fusion will involve mixed patient populations; we abstracted data from these studies only if outcomes are reported separately for individuals with chronic low back pain and otherwise uncomplicated DDD, or if >75% of patients carried such a diagnosis. Note that some surgical studies included patients who have attempted conservative management for varying lengths of time; these were included regardless of the duration and/or intensity of prior conservative management. Studies that include patients with a history of prior back surgery for any indication will be analyzed separately from patients undergoing lumbar fusion surgery for the first time.

Intervention
We evaluated the effectiveness of the major technical approaches to lumbar fusion surgery, regardless of surgical technique (e.g., anatomic approach, laparoscopic vs. open) or type of hardware utilized.

Comparators
Given open questions around the benefits of lumbar fusion versus nonsurgical management, we identified conservative management approaches as the primary comparator for this assessment. Conservative management options include physical therapy, intensive exercise/rehabilitation, cognitive behavioral therapy, and medication management, each alone or in combination. Other comparators of interest included minimally-invasive treatments (e.g., radiofrequency ablation, electrothermal therapy), if available. However, studies comparing lumbar fusion to artificial disc replacement were excluded, as artificial discs represent a separate review topic for the HCA.

Outcomes
Outcomes of interest included: 1) specified patient- and clinician-reported measures of pain, function, and disability; 2) opioid medication use; 3) requirements for repeat surgery or other retreatment according to type of initial surgery; 4) return to work and/or resumption of normal activities; 5) mortality, stratified according to cause of death where available; 6) other complications and adverse events; 7) measures of "treatment success" or "successful clinical outcome" (e.g., return to work and/or functional goals, cessation of pain medication, available composite measures); and 7) the total costs and cost-effectiveness associated with fusion in comparison to alternative treatment approaches. Functional status was recorded as measured by standard indices (e.g., Oswestry Disability Index [ODI][87], Roland-Morris Disability Questionnaire [RDQ][88]), back pain was recorded as measured by a visual analog scale (VAS), and health-related quality of life (HRQoL) was abstracted based on validated instruments (e.g., short-form [SF]-36 questionnaire). Of particular interest to this evaluation was measurement of treatment effects in comparison to varying intensities of conservative management (e.g., interdisciplinary rehabilitation vs. physical and/or behavioral therapy alone).

Recommendations from influential clinical societies and other authoritative sources inform interpretation of meaningful improvement as reported on validated measures for pain and/or function. For example, a mean 10-20 point change on a 100-point visual analog pain scale or 5-10 points on the RDQ are generally considered moderate improvements.[89] Other published thresholds for clinically-meaningful improvement include at a 30% decrease from baseline on a chronic pain scale or an improvement of at least 20 points on the ODI.[44] Importantly, while we sought data on these specific thresholds as reported in clinical studies, we abstracted all measures of clinically-meaningful change or treatment success as defined in each study, even if they differed from published guidance.

Information on the costs and cost-effectiveness of lumbar fusion procedures compared to alternative treatment was assessed using evidence from the available economic literature, including treatment-related costs, costs of long-term care (e.g., treatment switching, repeat surgery, complications, etc.), and indirect costs (e.g., productivity loss, caregiver burden).

Sources: Timeframe and Study Designs
Data on outcomes of interest were abstracted at all relevant timepoints. However, while perioperative benefits and risks of surgery (i.e., within 30 days) were of interest, so too were duration of benefit and other potential long-term effects. Because of this latter concern, we focused attention on longer-term comparative studies and/or timepoints in which at least 80% of the original sample was present.

We included randomized controlled trials (RCTs) as well as comparative observational studies without restrictions on study design parameters other than that there be explicit prospective or retrospective comparisons of at least one surgical procedure of interest to a non-surgical intervention.

Our primary focus of attention was on good- or fair-quality RCTs and comparative observational studies. However, for completeness, we abstracted data from case series meeting the following criteria: (a) sample size ≥100, (b) minimum follow-up of two years, (c) ≥80% patient retention, and (d) ≥75% with uncomplicated DDD or findings stratified by indication for fusion.

Literature Search and Retrieval

The PICOS were operationalized in the form of search strategies constructed for each of the literature databases used as a source of information, and as well in the form of inclusion/exclusion criteria for application to the publications identified through implementation of the search strategy.

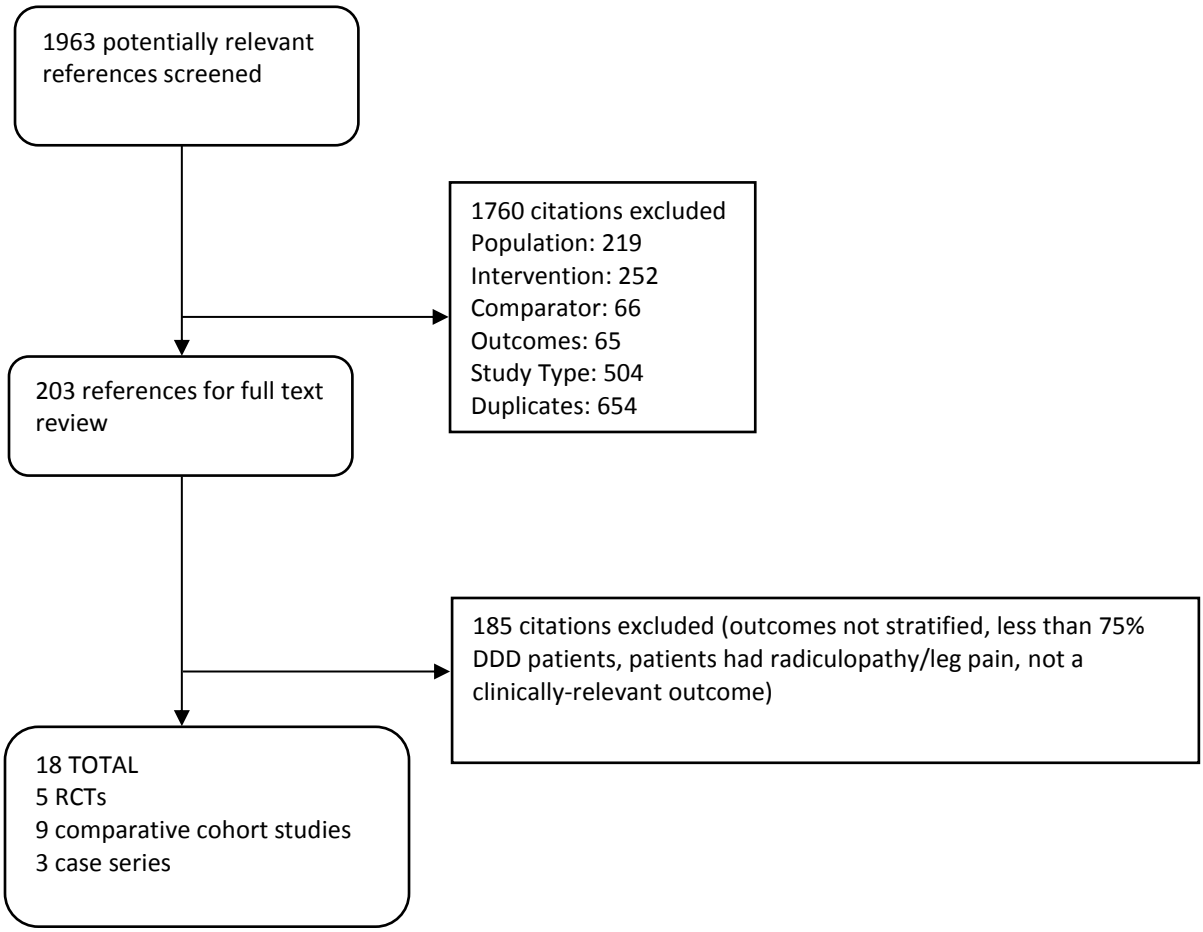
The timeframe spanned the period from January 2000 to the most recently published data available as of June 1, 2015 in the following electronic databases: MEDLINE (accessed through OVID), Cochrane Register of Controlled Trials, Databases of Abstracts of Reviews of Effects (DARE), OT Seeker, PEDro, ABI Inform, EconLit, and Health and Psychosocial Instruments.

Reference lists of all eligible studies were also searched and cross-referenced against public comments received by the HCA. Electronic searches were supplemented by manual review of retrieved references, previously published technology assessments, and systematic reviews. Further details on the literature search strategy can be found in Appendix A.

A single investigator screened titles and abstracts of all publications identified in the literature search, applying exclusion criteria if explicitly clear. A subset of excluded studies were reviewed by a second investigator as a quality control measure. The full text of all publications remaining after review of the titles and abstracts were retrieved, and the inclusion and exclusion criteria were applied to this set. As before, a subset of excluded studies were reviewed by a second investigator as a quality control measure.

The combined search results identified 1,963 potentially relevant studies for this assessment (Figure 2 on the following page). After elimination of duplicate and non-relevant references, we identified five randomized control trials, nine comparative cohort studies, and three case series, for a total of 18 included studies.

Figure 2: PRISMA flow chart showing results of literature search



Study Quality

Assessment of the quality of clinical trial reports and systematic reviews followed methods adapted specifically for studies of low back pain from the Cochrane Back Review Group.[17] For observational studies, we used the approach of the U.S. Preventive Services Task Force (see detailed descriptions on the following page.[18] Finally, while there are no published criteria for evaluating quality of case series due to their noncomparative nature, we identified specific quality criteria for inclusion of these studies as described above.

***Good**: Meets all criteria: Comparable groups are assembled initially and maintained throughout the study (follow-up at least 80 percent); reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; all important outcomes are considered; and appropriate attention paid to confounders in analysis. In addition, for RCTs, intention to treat analysis is used. Specifically for this review, target or mean/median duration of follow-up did not appreciably differ within study groups.*

***Fair**: Studies will be graded "fair" if any or all of the following problems occur, without the fatal flaws noted in the "poor" category below: Generally comparable groups are assembled initially but some question remains whether some (although not major) differences occurred with follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are addressed. Intention to treat analysis is done for RCTs. Specifically for this review, differences in baseline characteristics and/or duration of follow-up were allowed only if appropriate statistical methods were used to control for these differences (e.g., multiple regression, survival analysis).*

***Poor**: Studies will be graded "poor" if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. For RCTs, intention to treat analysis is lacking.*

Overall strength of evidence for each key question was described as “high”, “moderate”, or “low”, and utilized the evidence domains employed in the AHRQ approach.[18] In keeping with standards set by the Washington HCA, however, assignment of strength of evidence focused primarily on study quality, quantity of available studies, and consistency of findings.

In addition, summary ratings of the comparative clinical effectiveness and comparative value of the procedures of interest (i.e., *across* multiple key questions) were assigned using ICER’s integrated evidence rating matrix.[19] The matrix has been employed in previous Washington HCA assessments of virtual colonoscopy, coronary CT angiography, cervical fusion surgery, cardiac nuclear imaging, proton bean therapy, breast imaging in special populations, and bariatric surgery. The matrix can be found in Appendix C to this document.

Data Synthesis

Data on study design, population, and relevant outcomes were abstracted by a single reviewer, with additional review by a second review as a quality control measure. Qualitative evidence tables for the studies selected for review can be found in Appendix B. The findings were summarized descriptively as responses to each of the key questions to which this report is responding.

9. Results

9.1 Overall Evidence Quality

There were a number of specific limitations affecting the quality of the studies in the evidence base. Among these was an imbalance in treatment groups with respect to factors potentially influencing outcomes, or a lack of consideration of such factors in the analysis of the resulting data. Often, but not always, such imbalances were addressed by authors in the analysis phase of the study, presenting treatment effect estimates adjusted for the factors of concern.

Also of concern was the lack of longer-term follow-up data in many studies, and the lack of strict criteria defining treatment groups. Many study populations were subject to substantial attrition rates, limiting the power of such studies to document effect sizes of interest at these timepoints of interest. Additionally, treatment group definition was often heterogeneous. This precludes easy synthesis of findings with respect to both surgical and non-surgical interventions.

Of the five RCTs identified for this review, we rated three[20,22-24] (60%) to be of good quality based on the comparability of groups with respect to both baseline characteristics and duration of follow-up, and minimized sample attrition; and two[20,25] RCTs (40%)were rated as of fair quality. Quality issues affecting the RCTs are described in detail below. Six[26-29,31,33,34] of the eight prospective cohort studies were rated as good quality (75%), one[32] as fair (12.5%), and one[30] as poor (12.5%). A retrospective cohort study[34] was rated as poor. The poor quality ratings reflect the presence of at least one key quality issue not adequately addressed in either the design or analysis phase of the study.

In the study by Fairbank et al.[24], there was a substantial degree of crossover, with over 25% of patients randomized to intensive conservative management having had surgery by the end of two years; this is in contrast to only 4% of those randomized to surgery who crossed over to conservative management. A separate multiple imputation analysis was conducted to carry forward values for patients who crossed over or were lost to follow-up; this did not materially affect any primary findings. This study also described substantial imbalances between treatment groups in several potentially important baseline characteristics; as with the issue with crossover, the authors addressed this issue in the analysis phase, in this case by estimating the relative treatment effects in multivariate analyses controlling for these factors as additional independent variables.

In contrast to the Fairbank study, crossover rates in either direction between the group randomized to spinal fusion and the group randomized to non-intensive conservative management were relatively low (<10%) in the RCT by Fritzell et al.[25,38], and these crossovers were analyzed separately. However, the authors of this study failed to address any imbalances between the treatment groups with respect to factors possibly impacting treatment outcome; imbalances included mean pain duration between the groups and the presence of comorbidity. An additional limitation of this study included the lack of definition around conservative treatment. These limitations were not severe, but because no effort was made to evaluate their impact, the quality of this study was graded as fair, rather than good.

Two RCTs by Brox et al., were limited by small sample size despite the incorporation of a power calculation in the study design (total sample n=60[22] and n=64[23] in the 2003 and 2006 studies, respectively.) Both studies also had one year of follow-up, somewhat limiting the applicability of the

evidence to questions regarding the duration of treatment effect. These limitations were deemed minimal enough to support a quality rating of good for both studies.

The RCT described by Ohtori et al.[20] was also limited by sample size (total sample, n=41), and further by the lack of consistency in the type of fusion surgery performed in the surgical treatment group. These limitations downgraded the quality rating for this study to fair.

Key Question #1: What is the comparative clinical effectiveness of lumbar fusion surgery for patients with chronic low back pain and uncomplicated DDD relative to that of conservative management, minimally-invasive treatments, and other nonsurgical approaches?

We identified three good-quality RCTs, two fair-quality RCTs, four good- or fair-quality longer-term follow-up reports on these RCTs, one fair-quality secondary analysis, one good-quality prospective cohort study, and one poor-quality retrospective cohort study (see Appendix B for study details). Of note, none of these studies compared lumbar fusion to minimally-invasive treatments alone, and conservative management approaches varied across studies. Comparisons are further complicated by differences in study design, methods, and crossover rates. Based on the available evidence, lumbar fusion provides some advantage over lower-intensity conservative approaches (e.g., physical therapy or exercise alone) in improving pain and disability and returning to work over a shorter duration of follow-up (i.e., up to two years); however, differences diminish and are no longer statistically significant over longer durations of follow-up. Conversely, comparisons of lumbar fusion to more intensive and/or interdisciplinary forms of rehabilitation yield no differences in effectiveness.

We identified five RCTs[20,22-25] comparing lumbar fusion to conservative treatment among patients with uncomplicated DDD. Four of these studies[22-25] were evaluated in the original assessment[39] for the HCA; only one additional RCT[20] conducted in Japan was identified for this re-review. Three of these studies[22-24] compared fusion to interdisciplinary rehabilitation with a cognitive-behavioral component. The remaining two RCTs compared fusion to non-intensive physical therapy[25], or an exercise treatment plan[20]. While patients undergoing lumbar fusion had similar absolute levels of improvement in pain and function over one to two years of follow-up across four of the five RCTs[22-25], statistically-significant treatment effects favoring fusion were noted only in the RCTs comparing fusion to less intensive treatment. None of these RCTs included patients who had previously undergone fusion surgery, though three[22,24,25] allowed individuals with who had a prior discectomy.

Table 2 on the following page lists the study details of these five key RCTs. Several recent systematic reviews[17,40-42] evaluating these studies have noted that patient inclusion criteria and control treatment regimens may affect outcomes in a substantive way; more details on the effect of the treatment intensity in the conservative cohorts are reported in Key Question #4. The section below describes the short- and longer-term outcomes from these RCTs, as well as the nonrandomized comparative studies we identified as part of our literature search. The rate of harms associated with lumbar fusion versus conservative care are discussed in detail in Key Question #3.

Table 2. Study details for 5 key RCTs comparing fusion to conservative treatment in patients with uncomplicated DDD.

Study (Country of Origin)	Quality	Sample Size	Setting Type	Entry Criteria	Patient Characteristics	Control Group Description	Fusion Group Description	Follow-up Duration
Brox 2003[23] (Norway)	Good	64	Multicenter	Aged 25-60 CLBP ≥1 year *Patients who had undergone previous spinal surgery were excluded*	Age: 43 Pain duration: 10.8 years % male: 39	Cognitive intervention and individual exercises with increasing intensity	Posterolateral fusion with instrumentation and postoperative physiotherapy	1 year
Brox 2006[22] (Norway)	Good	60	Multicenter	Aged 25-60 CLBP ≥1 year *All patients had prior discectomy for disc herniation*	Age: 43 Pain duration: 8.0 years % male: 52 % prior discectomy: 100	Cognitive intervention and individual exercises with increasing intensity	Posterolateral fusion with instrumentation and postoperative physiotherapy	1 year
Fritzell 2001[25] (Sweden)	Fair	294	Multicenter	Aged 25-65 CLPB ≥2 years *Patients with successful discectomy >2 years before fusion were allowed*	Age: 43 Pain duration: 8.0 years % male: 49 % prior discectomy: 18.8	Non-intensive physical therapy + information and education aimed at pain relief	Noninstrumented posterolateral, instrumented posterolateral, or instrumented circumferential	2 years
Fairbank 2005[24] (UK)	Good	349	Multicenter	Aged 18-55 CLPB ≥1 year *Candidates for surgery irrespective of previous root decompression or discectomy*	Age: means reported by age groups Pain duration: 8.0 years % male: 49 % prior discectomy: NR	75 hours of IRP, including daily muscle strengthening and exercise, CBT, and hydrotherapy	At the discretion of the surgeon	2 years
Ohtori 2011[20] (Japan)	Fair	41	Single center	CLPB ≥2 years *Patients who had undergone previous spinal surgery were excluded*	Age: 34 Pain duration: 7.3 years % male: 59	Exercise treatment, including 30 minutes of daily walking and muscle strengthening	Anterior interbody fusion or posterolateral fusion with pedicle screws	2 years

Pain and Function

RCT-based evidence on lumbar fusion surgery versus intensive rehabilitation with a cognitive element comes from three studies[22-24] conducted in Norway and the UK. In the Norwegian RCTs[22,23], no significant differences were observed for pain (as measured by a 100-point VAS scale) or the ODI at 1 year of follow-up; medication use was also not significantly different in either study. Notably, in the later study[22] which included only those patients who had a prior discectomy, absolute changes on the ODI were nominally in favor of the conservative cohort (12.8 vs. 8.9 for surgery). Both studies reported a 97% follow-up rate, with only 2.4% of patients across studies switching to the surgical group after randomization.

Although a significant difference in the ODI favoring lumbar fusion was observed in the UK RCT[24] (-12.5 vs. -8.7, p=0.045) relative to IRP, the authors noted that this difference was only marginally significant. No significant treatment effects were noted for improvements on a shuttle walking test or any of the SF-36 subdomains or component summary scores. These results are confounded by differences between groups for follow-up at two years (78% and 84% in the surgical and conservative groups, respectively), with 28% of patients crossing over to the surgery compared to only 4% switching to the rehabilitation group. However, a separate multiple imputation analysis was conducted to carry forward values for patients who crossed over or were lost to follow-up; this did not materially affect any primary findings.

In the Swedish RCT[25], significant differences favoring surgery were observed in the mean change from baseline to year 2 for both the 100-point VAS (-21.0 vs. -4.3, p=0.0002) and the ODI (-11.6 vs. -2.8, p=0.015) relative to non-intensive physical therapy. However, after six months of treatment the benefits of surgery began to diminish, and the authors observed that back pain increased significantly between one and two years of follow-up for the fusion cohort (p<0.0001). Although this RCT had low attrition with only 2% lost to follow-up, crossover was noted in both groups, including 25% of patients in the rehabilitation cohort and 3% in the surgical group. In the most recent RCT[20] from Japan, there were statistically-significant improvements in favor of ABF and PLF versus exercise treatment (-51.7 and -44.8 vs. -24.0), VAS (-6.1 and -4.0 vs. -3.0), and JOA (+1.4 and +1.3 vs. +0.5) for ABF, PLF, and exercise treatment, respectively, over two years of follow-up (all outcomes, p<0.01). No patients were reported being lost to follow-up, or switching to a different treatment group. However, this small study[20] was largely focused on comparing differences between the two fusion techniques[43], and the control group was only "minimally-treated" with 30 minutes of physician-supervised daily exercises and stretching.

In addition to the above-described RCTs, good-quality follow-up data were available for three of the five RCTs. In a combined study[26] of the original Norwegian RCT cohorts[22,23] (n=124, mean age 43, 45.2% male) after a mean follow-up of four years (with 89% of the original population remaining), the adjusted treatment effect between fusion and non-operative care was non-significant. After nine years[28], patients from both groups (n=99, mean age 43, 38.6% male) who consented to long-term radiography follow-up had similar ODI scores. In a sensitivity analysis which included one-third of patients who crossed over to the surgery group, there were significantly more patients taking opioids on a daily or weekly basis in the surgical cohort compared to non-operated patients (44% vs. 17%, adjusted OR: 4.0; 95% CI: 1.5, 11.0; p=0.005), though no differences were observed in the intent-to-treat analysis. Another fair-quality follow-up study[31] with 261 patients (mean age 42, 47.5% male) pooled from the Brox[22,23] and Fairbank RCTs[24] also found no significant differences between groups on the ODI or VAS, as well as for pain medication use after a mean of 11.4 years of follow-up.

In addition to RCT data, we found one large, good-quality prospective cohort study[33] of 495 patients (mean age 43, 47.5% male) comparing surgery (79% instrumented fusion) to conservative treatment.

No specific treatment regimen was prescribed to either patient group in this observational study; rather, patients who were diagnosed with discogenic pain and received surgery within six months were considered part of the surgical group, and all others meeting the inclusion criteria were part of the non-operative cohort. Although the surgical group showed statistically-significant improvements over conventional treatment on the RDQ (-8.8 vs. -1.8) and SF-36 (PCS: +14.5 vs. +2.4) after one year (both outcomes $p<0.001$), the authors noted that the conservative group was minimally-treated, with only 5% receiving CBT, and is likely biased in favor of surgery due to patient selection. Opioid pain medication use was also not statistically-different between groups.

The final study[34] we identified as part of our literature search was a poor-quality retrospective cohort study (n=96, mean age 47, 50% male) comparing lumbar fusion to conservative treatment, which included physical therapy, epidural injections, and medication. This study did not find any significant differences between groups for Numerical Rating Scale pain scores, or the ODI after five years of follow-up. However, there are some substantial methodological concerns with this study, including the failure to control for significant differences in patient characteristics between individuals at baseline and those lost to follow-up, which was more than half of the original population.

Quality of Life

Data regarding the impact of lumbar fusion on quality of life were available from the Fairbank RCT[24], as well as the follow-up study[31] of the Fairbank and Brox RCTs. In the original Fairbank RCT[24], no statistically-significant differences were noted at 24 months for the SF-36 mental or physical component summary scores, nor were differences observed in any specific subdomain. In the long-term follow-up study of Fairbank and Brox[31], there were no significant differences between groups on the EQ-5D VAS for health-related quality of life (HRQoL) in both the intent-to-treat and as-treated analyses.

The above-described poor-quality retrospective study[34] found that HRQoL scores based on the ODI, SF-12 MCS, and SF-12 PCS were not statistically-different between groups.

Patient Satisfaction

Five good-quality studies reported information on patient satisfaction, but used varying definitions. Life satisfaction following treatment was rated on a 10-point VAS scale in the first Brox study[23]; there were no significant differences between groups after one year, nor were there any differences in the four-year[26], nine-year[26], or 11-year[31] follow-up studies for the Brox[22,23] and Fairbank[24] pooled cohorts. The Ohtori RCT[20] asked patients to state if their assigned treatment met their expectations according to criteria adopted from the North American Spine Society Low Back Outcome Instrument. After two years, 15 surgical and 10 non-operative patients voted that their treatment met their expectations, while two and six in the fusion and conventional management treatment groups, respectively, reported they were the same or worse after treatment. Results were not statistically-significant, although this small study was likely underpowered to detect differences between groups.

The above-described poor-quality retrospective study[34] also measured patient satisfaction based on a study-specific four-point scale from “very satisfied” (1) to “very dissatisfied” (4), but there were no statistically-significant differences between the operative and non-operative groups over a mean follow-up of five years.

Return to Work

Data on the impact of lumbar fusion on return to work come from the Norwegian and Swedish RCTs, and their subsequent follow-up studies. In first Brox study[23], the percentage of employed individuals

who returned to work was numerically higher in the intensive rehabilitation control group, but did not reach statistical significance. The 2006 study[22], which evaluated patients with prior disc herniation surgery, similarly found that although there were more patients from the intensive rehabilitation group working full-time, these numbers were too small to be evaluated statistically. In the pooled four-year[26] and 11-year follow-up studies[31], these differences continued to be non-significant.

In contrast, the percentage of patients in the Fritzell RCT[25] not working at baseline due to back pain who were employed at the end of the study was statistically-significantly in favor of the lumbar fusion group (39% vs. 23% for physical therapy, p=0.049). The "net" rate of back to work (i.e., subtracting those who stopped working during follow-up) was also significantly higher in the fusion group (36% vs. 13% for physical therapy, p=0.002). A subanalysis[29] of the original RCT found that a shorter duration of sick leave prior to treatment was significantly associated with work status at follow-up in both the surgical (14 months for those working, and 31 months for those not working, p<0.0001) and conservative (13 months for those working, and 27 months for those not working, p=0.006) groups. Other variables, including sociodemographics (e.g., gender, smoking, comorbidity), pain (e.g., duration of pain, quality of pain), clinical findings (e.g., reflexes, sensation), psychological diagnosis (e.g., personality disorders), or radiography (e.g., Modic sign type 1), were not significantly associated with work status at follow-up.

Mental Health

The most frequently-reported outcome beyond those described above was depression. Of the previously-described studies, two RCTs[24,25], one secondary analysis[29], and one prospective comparative cohort[33] evaluated differences between surgical and non-surgical cohorts for changes on depression scales, including the Zung Depression Scale (ZDS)[24,25,29] and the standard checklist-90 (SCL-90)[33]. Neither RCT[24,25] found any significant differences between groups for depression. In the sub analysis of the Fritzell RCT[29], however, patients in the conservative group were significantly more depressed than the fusion cohort after two years of follow-up (31 vs. 37, p<0.0001). It is worth noting that the ZDS was modified from a 20-80-point scale, to a 0-100-point scale (where 100 represents maximal depression) for this study to capture "psychological distress," which may have influenced outcomes. The authors also reported that higher depressive symptoms at baseline were predictive of improvement for patients in the conservative group (OR 1.08, 95% CI: 1.02, 1.14); this effect was not significant in the fusion cohort, however.

We identified only one observational study[33] in our search that evaluated depression as an outcome. Although there were no statistical differences between groups for up to six months after treatment, patients in the fusion group were significantly less depressed than those receiving unstructured non-operative care after nine (+0.16 vs. -0.4, p=0.029) and 12 (0.0 vs. -0.4, p<0.001) months of follow-up from baseline based on a 0-4 depression scale.

Key Question #2: What are the rates of "treatment success" or "successful clinical outcome" of lumbar fusion as defined by measures of clinically-meaningful improvement in pain, function, quality of life, patient satisfaction, and/or work status?

Much of the work done to quantify clinically-significant improvement in measures of pain and function at the individual patient level came after the publication of the RCTs of interest. Two of the five RCTs we identified for this assessment did not include any measurement of "successful" outcome. Findings from the other three RCTs, as well as one prospective cohort study, mirrored those of continuous

measures of effectiveness in that results favoring surgery were limited to studies that compared surgery to minimal or nonspecific approaches to conservative management.

In recent years, multiple efforts have been undertaken to identify clinically-meaningful changes in measures at the individual patient level. These individual "success" outcome measures include a mean 10-20 point change on a 100-point visual analog pain scale or 5-10 points on the RDQ, which are generally considered moderate improvements.[17] Other published thresholds for clinically-meaningful improvement include at least a 30% decrease from baseline on a chronic pain scale or an improvement of at least 20 points on the ODI.[44] Patient-defined minimum acceptable outcomes also include discontinuation of opioid medication and return to some occupational activity, though individuals with significant psychosocial factors (e.g., compensation claims, psychological distress), may be less likely to report satisfaction with treatment despite achieving the desired outcomes.[45]

Unfortunately, the development of measures of clinically-meaningful change at the individual level came after publication of all but the small Ohtori[20] RCT. Measures of treatment success in earlier RCTs were limited to patient-reported or independent observer assessment of improvement after intervention. In the Fritzell RCT[25] comparing fusion to physical therapy of varying intensity, 63% of patients in the surgical group rated their symptoms as "much better" or "better" compared to 29% receiving conservative management ($p<0.0001$). Results were rated as "excellent" or "good" by independent observers for 45% and 18% of patients in the surgical and conservative groups, respectively ($p=0.005$). In contrast, there were no statistically-significant differences in either patient or independent observer ratings of treatment success in the two Brox[22,23] RCTs comparing fusion to cognitive/exercise intervention. Measures of treatment success were not considered in either the Fairbank[24] or the Ohtori[20] RCTs.

Some of the studies include mention of clinically-meaningful change in their Discussion sections. Fairbank[23] and Brox[23] (2003) remark that the mean difference in ODI scores between groups did not approach 10.0, which was considered a clinically-meaningful difference. In fact, the confidence interval in the Fairbank RCT did not include 10.0, essentially ruling out any possible difference in favor of surgery. In the Brox 2006[22] RCT, the observed mean difference in ODI after adjustment for gender and pretreatment expectations was 9.7 points, and the confidence interval around this result included the possibility that exercise/cognitive therapy was *superior* to fusion.

Recent nonrandomized studies have made use of published measures of clinically-meaningful improvement, but their number is extremely limited for patients with uncomplicated DDD. A single good-quality prospective cohort study[33] evaluated clinically-meaningful improvement between treatment groups based on a 30% or 5-point improvement on the Roland-Morris back disability score and found that, after controlling for baseline differences, surgery was significantly better than conservative treatment based on this criteria (57% vs. 25%, $p<0.001$). In addition, 33% and 15% of patients in the surgical and conservative groups achieved a composite measure of treatment success that included the above Roland-Morris thresholds as well as a ≥30% improvement in pain intensity, no use of opioid pain medication, and a status of employed at 12 months ($p<0.001$). While these results favored surgery, the authors cautioned that the control group received a variety of interventions and overall, did not appear to receive services consistent with major guidelines for treatment of chronic low back pain. For example, only half of patients received any physical therapy and 5% received a cognitive-behavioral intervention.

Only one case series that met our study inclusion criteria assessed a clinically-meaningful threshold of specific outcome measures for patients undergoing lumbar fusion surgery for uncomplicated DDD. Anderson et al.[46] prospectively evaluated 106 patients who received fusion (ALIF technique with titanium cages and autogenous iliac bone graft) and found that patients who were employed before surgery were significantly more likely to be working after a mean 29.7 months of follow-up (92% vs. 43%, OR 10.5, p=0.0008). An attempt to identify predictors of achieving 30% improvement on the RDQ using multivariate logistic regression found no statistically-significant associations between this outcome and work status, age, smoking history, gender, worker's compensation status, pre-operative pain or RDQ scores, and type of fusion surgery.

Key Question #3: What are the rates of adverse events and other potential harms (perioperative, long-term adverse events, and reoperations) associated with lumbar fusion surgery compared to alternative treatment approaches?

Evidence on harms in published RCTs of treatments for patients with chronic low back pain and uncomplicated DDD is limited by several factors. Many of these studies are too small to capture reliable data on complications that occur infrequently, and the relatively low rate of serious complications has led to standards for research reporting that often do not include a formal assessment of all complications. Other factors contributing to the dearth of data on harms include the lack of observational studies that focus on uncomplicated DDD patients, and the short-term nature of many studies, leading to a failure to observe adverse outcomes associated with surgical interventions that do not manifest until later years (e.g., repeat surgery). Harms associated with conservative treatment are rarely reported and are generally limited to non-compliance with the treatment protocol.

Unlike findings for clinical effectiveness, harms data are often not stratified for interventions that are used for multiple indications (e.g., both uncomplicated DDD and more specific indications). Rather than look to studies comparing different technical approaches of lumbar fusion, which are subject to the same methodological concerns as studies with a non-surgical comparator group (e.g., small sample sizes, shorter duration of follow-up, lack of standardized reporting), we have identified several large database studies evaluating harms associated with lumbar fusion across several indications to provide additional context on the rate of adverse events. These data are evaluated separately from our study set because either the majority of patients did not have a primary indication of uncomplicated DDD, or outcomes were not stratified for this population.

Lumbar Fusion

For lumbar fusion procedures, we have categorized harms as surgery-related mortality, overall adverse events (as reported in the included studies), and requirements for retreatment (e.g., reoperation/revision surgery). Although these studies used various technical approaches to fusion, we did not make any attempt to stratify outcomes by surgical method. Such data, if available, are summarized for Key Question #4.

Mortality

No data on perioperative mortality attributable to lumbar fusion were reported in any systematic review, RCT, or observational study that met our inclusion criteria. Overall mortality was reported in the Mannion[31] study; 7.1% (10/140) patients died in the fusion group and 0.8% (1/121) patients died in the conservative treatment group during the 11-year follow-up period for the Brox[22,23] and Fairbank[24]

cohorts. The authors noted that they could not definitively determine if these deaths were associated with chronic low back pain or its treatment given that some patients had illnesses unrelated to back pain, nor was this difference statistically tested.

Adverse Events

The most frequently-reported adverse events occurred during the perioperative period and included dural tears, bleeding, and wound infection, occurring at a rate of 9-18% in available RCTs and observational studies. Notably, the only RCT published since the original review[20] did not evaluate the rate of complications in either treatment group.

In the Fairbank RCT[24], a total of 19 patients experienced complications from surgery (10.8%), which were primarily dural tears and problems with surgical implants (2.8% each). In the 2003 Brox RCT[23], complications included two wound infections, two bleedings, one dural tear, and one venous thrombosis. Overall 6 patients (18.2%) experienced a complication, and all presented as early complications; there were no late complications associated with surgery. The 2006 Brox RCT[22] reported wound complications in only two patients (8.7%). During long-term follow-up for these studies[26], no additional complications related to surgery were reported. Fritzell et al.[25] reported 53 early complications occurring in 17% of patients, and 13 (6%) of patients suffered a late complication (defined as more than two weeks after surgery), including 9 patients who developed nerve root pain related to the pedicle screw implant. Overall, there were 16 (7.8%) unintended reoperations related to compilations in the fusion cohort.

We identified only one small, poor-quality prospective comparative cohort study[30] which evaluated outcomes for patients with degenerative disc disease (n=46, mean age 55, 59% male) undergoing minimally-invasive fusion surgery compared to those who had a previous discectomy undergoing fusion for the first time. Although more patients in the revision group experienced dural tears, overall there were no statistically-significant differences in perioperative complications between the groups.

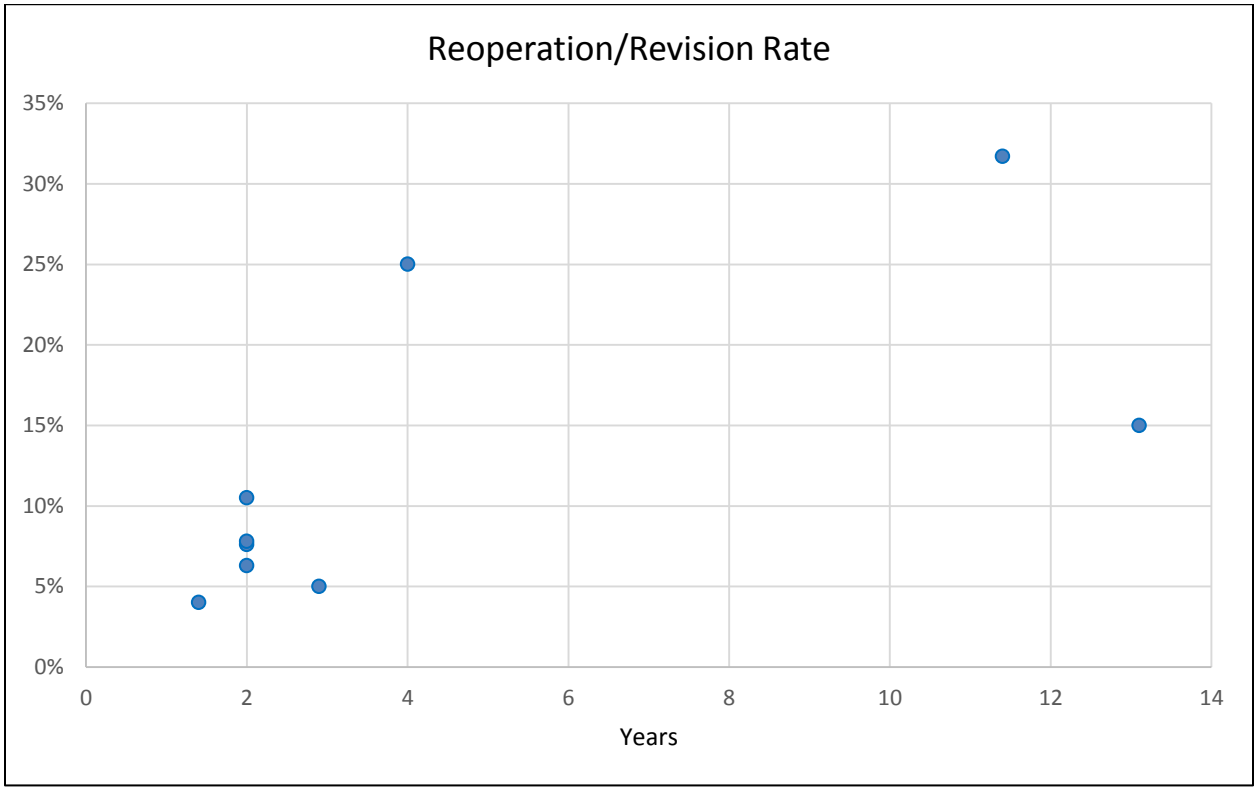
One case series[36] of 118 patients did not reported any intraoperative or major complications after surgery, but 2 patients (<1%) experienced a hematoma and one patient received a permanent disability rating. Complications rates in case series tend to be lower than in RCTs and cohort studies, which is not surprising given the information biases attendant in evaluations.

Subsequent Treatment

Data from available studies indicate that requirements for additional surgery vary widely in both reported rate and indication for such surgery. Across all studies, the rate of reoperation and/or revision surgery averaged approximately 12.5% across studies over a mean of five years of follow-up. As shown in Figure 3 on the following page, reoperation continues to be a concern even years after initial surgery. Studies of shorter duration (i.e., up to two years) had a lower reported rate of reoperation (4%-11%) compared to the limited number of studies with longer follow-up periods (15%-32%). Indications for additional surgery include hardware removal, repeat fusion, alternative lumbar surgery (e.g., discectomy), or some combination. The figure on the following page represents those studies[24-26,28,30,31,33,36,37] in our set that reported on the rate of reoperations. It is difficult to distinguish between revision surgery and reoperations for two reasons: 1) studies often use these terms interchangeably, and 2) patients can undergo surgery for multiple indications (e.g., a combination of hardware removal and repeat fusion), so reasons for repeat surgery are not always stratified. One study[31] reported these outcomes separately; of the 38 (15%) patients requiring additional surgery, 17 involved hardware

removal, 11 required repeat fusion, nine had a combination procedure, and one underwent a discectomy.

Figure 3. Rate of reoperations/revision procedures across all studies reporting this outcome.



Interestingly, only one[36] of these studies associated repeat surgery with adjacent segment degeneration, which is considered a major concern with lumbar fusion[47] and can cause recurrent lumbar pain. Lammli and colleagues reported that one-third of the additional surgical procedures were performed due to degeneration adjacent to the primary fusion level. Two additional long-term studies in our sample evaluated this outcome but with conflicting results. Froholdt et al.[27] (n=48, mean age 43, 42.8% male) included patients from the Brox RCTs[22,23] who had radiographs available for review, and found no differences between the surgical and conservative groups after a mean of nine years follow-up. In contrast, another follow-up study[32] which included 369 patients (mean age 43, 46.7% male) who participated in the Brox[22,23], Fairbank[24], and Fritzell[25] RCTs who consented to long-term radiographic follow-up over a mean duration of 13.1 years found a significant correlation between surgery and adjacent segment degeneration by assessing adjacent disc height (TE: -0.44 standard deviations, 95% CI: -0.77, -0.11; p=0.01), but this correlation was not associated with statistically-significant changes in patient-reported measurements of pain or disability.

Conservative Care

Conservative treatment in the available studies was typically not subject to a specific protocol, and involved a variety of non-operative treatment, including medications, physical or exercise therapy, intensive rehabilitation, and cognitive interventions. No attempt has been made to systematically evaluate potential harms from studies focused specifically on conservative management modalities.

Mortality
No cases of 30-day or overall mortality attributable to conventional or non-operative care, including interdisciplinary rehabilitation, physical therapy, or exercise treatment, have been reported in any systematic review, RCT, or observational study that met our inclusion criteria.

Complications
There were no reported complications of conservative or non-operative care, including interdisciplinary rehabilitation, physical therapy, or exercise treatment, in any systematic review, RCT, or observational study that met our inclusion criteria.

Subsequent Treatment
The only subsequent treatment associated with the conventional or non-operative care group in any study was related to non-adherence to the treatment protocol (i.e., cross-over to surgical cohort) due to persistent complaints or exacerbation of symptoms, though these were not described in detail and not necessarily related to conservative treatment.

Large Database Studies of Lumbar Fusion

As mentioned previously, we identified six large database studies evaluating complications for fusion across several indications (e.g., stenosis, isthmic spondylolisthesis, scoliosis, etc.) that did not met our inclusion criteria but are described here to provide additional information on complications associated with lumbar fusion. Three studies[48,49,53] used the National Inpatient Sample (NIS) database, two[50,51] studies evaluated data from Washington State-specific databases, and one study[52] reviewed the Swedish Spine registry for the 2011 calendar year.

The most recent study[49] to use NIS data to evaluate three different primary interbody fusion cohorts (923,038 fusions) over nine years. Patients with uncomplicated DDD represented a majority of patients for each fusion group (80.1%, 60.6%, and 78.6% for anterior lumbar interbody fusion [ALIF], posterior/transforaminal lumbar interbody fusion [P/TLIF], and combined anterior-posterior interbody fusion [APF], respectively, mean across groups: 64.2%). Table 3 below represents the rate of complications among these groups, showing a significantly higher rate for APF for 12 of 16 complications, and a significantly higher rate of mortality for ALIF. These rates were not adjusted for differences in baseline characteristics.

Table 3. Comparison of complications among P/TLIF, ALIF, and APF[49].

Complications	ALIF (%)	P/TLIF (%)	APF (%)	p-Value
Mortality	**0.25**	0.15	0.18	<0.001
Dysphagia	**0.17**	0.13	0.11	0.0017
Device Related	**5.43**	2.44	3.89	<0.001
Neurologic	0.37	**0.96**	0.55	<0.001
Cardiac	0.90	0.87	**1.23**	<0.001
Peripheral Vascular	0.22	0.08	**0.28**	<0.001
Respiratory	1.65	1.25	**2.20**	<0.001
Gastrointestinal	4.83	2.20	**5.56**	<0.001
Genitourinary	0.84	1.02	**1.11**	<0.001
Postoperative Shock	0.08	0.08	**0.13**	0.0002
Hematoma/Seroma	0.63	0.62	**0.82**	<0.001
Intraoperative Accidental Puncture/Laceration of Nerve/ Blood vessel	3.41	3.43	**4.20**	<0.001
Wound Dehiscence	0.27	0.15	**0.37**	<0.001
Postop Infection	**0.74**	0.43	**0.74**	<0.001
Acute Anemia secondary to Hemorrhage	7.39	11.42	**11.60**	<0.001
Acute Respiratory Distress Syndrome	**1.36**	0.75	**1.36**	<0.001
Venous Thromboembolic Events	0.62	0.41	**0.73**	<0.001

Table key: ALIF, anterior lumbar interbody fusion; APF, anterior-posterior interbody fusion; ARDS, acute respiratory distress syndrome; CNS, central nervous system; GI, gastrointestinal; GU, genitourinary; P/TLIF, posterior/ transforaminal lumbar interbody fusion; VTEs, venous thromboembolic events.

Note: Highest percentage is given in **bold**. p Value is from chi-square test.

Those studies[48,53] that did not meet our inclusion criteria (primarily because they did not have a majority of patients with uncomplicated DDD or report outcomes specific to this population), but reviewed large samples from the NIS database, evaluated whether mortality was associated with the incidence of specific complications of lumbar fusion across multiple diagnoses. The first study[48] identified a sample of 220,522 patients who had a fusion procedure (ALF, PLF, or APLF) for degenerative diseases of the lumbar spine and found that the incidence of postoperative ileus was significantly higher in those who had ALF surgery relative to PLF surgery (4.9 vs. 26.0 per 1,000) . Presence of postoperative illeus was associated with significantly higher Charlson comorbidity index (CCI) scores (3.05 and 2.13 for PLF and ALPF, respectively, p<0.001), and rates of mortality in both the ALF (1.5 vs. 4.1 deaths per 1,000, p=0.025) and PLF (1.1 vs. 4.0 deaths per 1,000, p<0.001) fusion groups. The second study[53] evaluted the incidence and potential risk factors of cerebral vascular accidents (CVA) following lumbar fusion surgery. A total of 340 CVAs out of 264,891 fusions (1.3 per 1,000) were identified between 2002-2011, and were associated with a greater mortality rate (73.7 vs. 0.8 per 1,000 patients) compared to those who did not have a CVA. Risk factors associated with CVA include advanced age (64.4 vs. 55.0 years for no CVA) and preoperative comorbidies as demonstrated on the CCI (4.03 vs. 2.52 for no CVA) (both outcomes, p<0.001).

Two additional database studies[50,51] reviewed Washington state-specific data to identify complications and mortality associated with lumbar fusion procedues. One of these studies[50] used the Comprehensive Hospital Abstract Reporting System (CHARS) registry of all nonfederal hospitals in Washington State and identified 5,091 adults who underwent a primary fusion procedure for degenerative diseases of the lumbar spine between 2004 to 2007. The overall complication rate for patients with DDD (n=1,097 or 18% of the total population) within the first 90 days after surgery was 4.2%, 2.1% had a repeat lumbar fusion surgery, and there were no deaths. During the one year follow-up, an additional 3.2% had a reoperation, but no deaths or complications were observed. The second study[51] identified all workers' compensation claimants (n=2,378) who underwent fusion from 1994 through 2001 and found a 90-day perioperative mortality rate of 0.29% (95% CI: 0.11%, 0.60%) and a 3-year cumulative mortality rate of 1.93% (95% CI: 1.41, 2.57). Interestingly, patients without a specific indication for surgery were more likely to experience the adverse consequences of narcotic use; a diagnosis of degenerative disc disease was associated with the highest risk of analgesic-related mortality (Risk Ratio [RR] 2.71, 95% CI: 1.17, 6.28).

The final database study[52] retrospectively reviewed the Swedish National Spine Register from 2011. In a cohort of 3,066 patients who had fusion surgery, 14% underwent reoperations over a mean three years of follow-up, of which 53% were related to removal of an implant and 47% were related to other complications from surgery. A minority of patients (8%) were listed as having a sole diagnosis of DDD and 38% of patients had previous lumbar spinal surgery; however, no further details on complications were reported.

Key Question #4: What is the differential effectiveness and safety of lumbar fusion according to factors such as age, sex, race or ethnicity, pre-existing conditions (e.g., smoking history), intensity of conservative management (e.g., interdisciplinary rehabilitation vs. physical and/or behavioral therapy alone) technical approach to fusion (e.g., posterolateral vs. interbody, minimally-invasive vs. open procedures), initial vs. repeat surgery, insurance status (e.g., worker's compensation vs. other), and treatment setting (e.g., inpatient vs. ambulatory surgery)?

There is little evidence to suggest that greater surgical intensity is related to changes in outcome in the long-term; advantages to less-intensive surgery (e.g., the effect of minimally-invasive surgery compared to open surgery was positive on HRQoL[18]) were noted in the short-term but did not persist in longer-term follow-up >2 years. On the other hand, our review suggests that more intensive and interdisciplinary rehabilitation featuring behavioral intervention may be both superior to usual-care approaches featuring only physical or exercise therapy, and that these more intensive approaches produces comparable outcomes compared to lumbar fusion. Workers' compensation status appears to have a differential treatment impact, negatively affecting some surgical outcomes (but not those of conservative management). This impact on surgical outcomes was inconsistent, however, as were the impact of age and gender. Our review did not find smoking status or BMI to be predictive of surgical outcome. These findings suggest that it will be difficult to use such factors to define subgroups of patients with DDD in whom surgical or conservative interventions would be preferentially indicated.

There are scant and often conflicting data addressing intervention-associated and patient-based factors that may influence outcomes following treatment for uncomplicated DDD. Several factors (e.g., age,

gender, complexity of fusion) are often adjusted for in analysis of the effect of treatment for DDD on various outcomes of interest; however, the rationale for variable selection and/or results of stratified analyses suggesting differential effects are rarely provided.

The evidence on differential effects of lumbar fusion according to various patient- and treatment-defined subgroups is summarized in the sections that follow. We gave priority to evidence from comparative studies where available, but also used data from fusion case series to augment our analyses.

Intervention Intensity

There have been five major RCTs published comparing spinal fusion to non-operative care among patients with non-specific low back pain. Three of these studies[22-24] compared fusion to "intensive" interdisciplinary rehabilitation with a cognitive-behavioral component, while control therapy in the remaining two RCTs was less intensive, at the discretion of the treating physician and mainly involving non-intensive physical therapy in one[25] and exercise in another[20]. The results of these studies with respect to benefits and harms associated with treatment have been presented previously in responses to Key Questions 1 and 2, but are summarized in further detail below.

Surgical Intensity

Within the primary review scope, patients undergoing spinal fusion had similar levels of improvement in pain and function over one to two years of follow-up across all five identified RCTs comparing surgical to conservative treatment. However, statistically-significant treatment effects favoring fusion were noted only in the RCTs comparing non-intensive physical therapy or exercise to PLIF without decompression or ALIF or PLIF with or without variable screw placement.[25] This is in contrast to a lack of significant findings in RCTs comparing intensive conservative management strategies to PLIF with posterior transpedicular screws[25] or to a range of fusion options[24].

Our review did not identify any publications describing the impact of previous surgery on the relative effect of surgical intervention for uncomplicated DDD compared to intensive conservative therapy. Two RCTs reported no benefit of lumbar fusion over intensive conservative management among patients with previous surgery for disc herniation[22,23]; this finding mirrors the lack of benefit noted for lumbar fusion over intensive conservative management among patients with no previous surgery Additionally, a prospective study of minimally invasive TLIF performed in 25 patients as a primary surgical intervention and in 21 patients as a revision documented no pain or function differences between primary and revision surgery at 1 year; these findings support the observation that there are few differences in primary versus revision surgery among patients with uncomplicated DDD treated with a surgical intervention.

The impact of the level of fusion on relative treatment effect of fusion versus conservative management was not evaluated in the five RCTs identified in our review. Our review also identified one case series[46] describing outcomes in a population of 106 patients with discogenic back pain followed for a mean of 29.7 months after treatment with varying intensity of ALIF (according to level of arthrodesis). Using a multivariate regression model, the authors evaluated the impact of single versus multiple-level fusion on a number of different outcome measures: return to work, a 30% improvement in the VAS pain score, or an increase of at least 30% on the Roland Morris score.[46] Fusion level was not found to be statistically-significantly associated with any of these outcomes.[46] Outside the body of primary literature identified within the scope of this review, several reports offer additional information regarding the impact of

lumbar fusion of varying intensity. The impact of differing levels of fusion (1, 2, or 3 or more) was evaluated in a retrospective study of 143 active duty military personnel (mean age 36.3 years), of whom 118 (83%) were DDD patients, treated with TLIF and followed for a mean of 34.9 months.[37] The level of fusion was not associated with the likelihood medical separation (i.e., an inability to remain on active duty).[37]

The impact of minimally-invasive vs open surgery was evaluated in a systematic review reporting the impact of these two approaches to PLIF surgery.[18] The findings of this review suggest that minimally invasive techniques may be associated with better HRQoL outcomes in the short-term, though the effect was variable, and not present at all in longer term follow-up (>2 years).[18] Primary reports reflecting these findings include a prospective study of 66 patients undergoing single level TLIF, comparing those experiencing open (n=33, of which 14 were patients with DDD) versus minimally invasive surgery (n=33, of which 13 were patients with DDD),[54] there were significantly lower VAS pain scores at 6 months post-surgery among those treated with a minimally invasive approach; no longer term data were presented.[54] Likewise, retrospective study of 64 patients receiving either minimally invasive TLIF or open TLIF for the treatment of DDD or spondylolisthesis reported lower VAS pain scores in the early post-operative period for the minimally invasive treatment, with no longer term data presented.[55]

The impact of instrumentation in lumbar fusion surgery was evaluated in a retrospective analysis of 1,310 DDD patients undergoing lumbar fusion, examining the impact of varying levels of surgical instrumentation on HRQoL, pain and function, and return to work.[56] Patients undergoing non-instrumented fusion (n=115) had higher levels of pain as measured on a VAS scale than those undergoing instrumented interbody fusion (p=0.02), although no differences in either HRQoL (as measured using the EQ-5D) or disability (as measured using the ODI) were noted.[56] Another randomized trial of patients with DDD treated with PLF (n=72) vs PLIF (n=73) reported no ODI or VAS differences between the 2 groups at 36 months.[57] These findings were supported by a prospective study of patients with DDD treated with PLF (n=82) and PLIF (n=80), in which no difference between the 2 groups was noted for ODI.[90]

The impact of cage use in lumbar fusion surgery was evaluated in a recent systematic literature review, which reported that single cage lumbar interbody fusion had significantly lower rates of complications than did two-cage fusion surgery (OR 0.30 [0.10, 0.95]).[91] Supporting this finding are those of a retrospective population-based cohort study of 2,378 chronic back pain patients treated with lumbar fusion surgery and receiving Washington State workers compensation who were followed by for a mean of 6.6 years; in this study, the use of cages or instrumentation was associated with increased complication rate compared with bone-only fusion surgery (OR 2.20 [1.16, 4.16]), without any improvement in disability or reoperation rates.[60]

Surgical Approach

The primary focus of our review was on comparisons of lumbar fusion to non-operative management; we nevertheless summarize available data comparing different forms of fusion below, with a focus on uncomplicated DDD where possible.

There exists little conclusive evidence documenting the impact of surgical approach on the outcomes of lumbar fusion among patients with uncomplicated DDD. A five-year RCT comparing the clinical outcomes of posterior midline fusion (n=25) compared to a paraspinal approach (n=25) in DDD patients reported significant improvement in outcomes for both groups, but no differences between groups.[61] Another RCT with 2 years of follow-up reported no statistically significant differences in function (ODI) or

pain (VAS) between groups of DDD patients with radiculopathy treated with TLIF (n=51) and PLF (n=47).[62] Evaluating the hypothesis that APF, with its anterior approach, may result in a higher incidence of major complications than TLIF; a respective analysis of 68 DDD patients treated with APF compared to 65 with TLIF reported higher rates of intra-operative complications associated with APF, and higher rates of post-operative complications associated with TLIF, with similar clinical outcomes in both groups.[63] A retrospective database analysis similarly documented a significantly increased incidence of postoperative ileus ALF surgery compared to PLF surgery (74.9 vs. 26.0 per 1,000; p<0.001).[48] A prospective observational study documented two-year outcomes associated with posterior fusion with translaminar screw fixation compared to TLIF in a cohort of 120 patients with DDD, and reported no difference in either clinical outcomes or treatment satisfaction.[64]

Surgical Setting

Our review did not identify any publications describing the impact of inpatient versus outpatient surgery on the relative effect of surgical intervention for uncomplicated DDD compared to conservative therapy.

Conservative Management Intensity

Conservative management in the five identified RCTs[20,22-25] incorporated a range of options, and differed in intensity. Table 4 on the following page describes the various components of the conservative management programs in each of the RCTs identified.

Table 4. Components of Conservative Management Programs Incorporated as Comparators in RCTs Evaluating Lumbar Fusion in the Treatment of Uncomplicated DDD

Publication	Conservative Management Components						Program Intensity
	Comparator	Strength Training	Aerobic Exercise	Educational Interventions	Biopsychosocial Interventions	Other Interventions	
Brox, 2003[23]	PLIF		Individualized endurance and coordination exercises	Rehab specialist lecture -Daily reinforcement	Fear avoidance Belief modification		75 hours/3 weeks
Brox, 2006[22]	PLIF			Rehab specialist lecture -Daily reinforcement	Fear avoidance Belief modification		75 hours/3 weeks
Fairbank, 2005[24]	Various fusion	Muscle stretching Spinal flexibility General strength Spine stability	Individualized endurance and coordination exercises		CBT: Fear avoidance and belief modification	Hydrotherapy	60-110 hours/3 weeks
Fritzell, 2001[25]	PLF, ALIF, or PLIF*	Ad hoc physical therapy		Ad hoc educational programs	Ad hoc cognitive training		NR
Ohtori, 2011[20]	ALIF or PLIF*	½ hour daily muscle stretching	1 hour daily walking				1095 hours/2 years

*Statistically significant treatment effect of surgery over conservative management

The conservative management programs differ with respect to the intensity of the intervention, with three[22-24] programs providing intensive treatment over a period of less than one month, and another two[20,25] providing treatment either over a longer period of time or with an undefined intensity. While comparisons across these RCTs are complicated by differences in study design, methods, and crossover[40], there are discernable patterns. Patients undergoing spinal fusion had similar levels of improvement in pain and function over one to two years of follow-up across all five identified RCTs comparing surgical to conservative treatment outcomes. However, statistically-significant treatment effects favoring fusion were noted only in the two RCTs[20,25] comparing fusion to non-intensive physical therapy or exercise. However, there appears to be relative benefit conferred by intensive non-surgical management compared to surgery.[22-24] No particular component of the management programs appears to be substantially associated with a greater relative benefit compared to surgery; such greater relative benefit appears instead associated with structure and intensity of the program over the short-term perioperative period.

Our review did not identify any studies directly comparing conservative management programs of varying intensity. Outside the scope of our review, there is evidence describing the relative effectiveness of varying intensity of conservative management. Several RCTs describe the efficacy of intensive interdisciplinary rehabilitation programs compared to specific physical therapy regimens.[65,66] Findings from those RCTs comparing higher intensity conservative management to some form of physical therapy were consistent, in that no significant treatment effects favoring the more intensive program were observed for any primary outcome measure; substantial improvements in pain, disability, and function were observed in both treatment groups.[65,66] Several systematic reviews describing the effectiveness of higher intensity programs have also been published. One review[67] found that intensive interdisciplinary rehabilitation programs (>100 hours) were associated with clinically-important improvement in function compared to usual care, while another review[68] did not find such an association between program intensity and clinical benefit. In sum, there is moderate evidence that intensive conservative management programs confer some level of incremental benefit over usual care, but not necessarily over less intensive programs of physical therapy.

Sociodemographic Factors

Age

Our review identified three good quality studies evaluating age as a potential predictor of treatment outcome: one RCT[29] and two case series[35,46]. The RCT [29] is a secondary analysis of data derived from the Swedish Lumbar Spine Study as described above[25,38]. The authors found that working status at the end of the 2-year follow-up was associated with younger age (evaluated as a continuous variable) in the surgical treatment group, but not in the non-surgical group, indicating a differential impact of age on treatment.[29] Supporting the impact of age on return to work was another case-series identified by our review, of 620 patients with DDD treated with single level posterolateral fusion, followed at least 3 years, of whom 24.4% returned to work in within 2 years postoperatively.[35] Negative predictors of return to work included age more than 50 years at fusion (OR 0.66; 95% CI: 0.45, 0.95).

Our review also identified another good quality case-series with results contrasting with those above. This study described outcomes in a population of 106 patients with discogenic back pain treated with ALIF and followed for a mean of 29.7 months.[46] Using a multivariate regression model, the authors evaluated the impact of age on a number of different outcome measures: returning to work, a 30%

improvement in the VAS pain score, or an increase of at least 30% on the Roland Morris score.[46] Age was not found to be associated with any of these outcomes.

Outside of the scope of the current review, there are conflicting data around the relationship between age and the outcome of surgical treatment for uncomplicated DDD. A retrospective population-based cohort study of 2,378 chronic back pain patients treated with lumbar fusion surgery and receiving Washington State workers compensation followed by for a mean of 6.6 years reported that age greater than 30 was significantly associated with higher rates of work disability, to the greatest degree in the oldest age group, greater than 60 (OR 3.07 [1.71-5.51]) compared to the reference group (those below 30). In contrast to this finding, another retrospective study, of 143 active duty military personnel (mean age 36.3 years), of whom 118 (83%) were DDD patients, treated with TLIF and followed for a mean of 34.9 months, younger age was associated with medical separation (an inability to remain on active duty) (OR for each additional year of age 0.93 [0.87, 0.98], p=0.01).[37]

Gender

Our review identified two publications describing a single good quality RCT evaluating the effects of fusion with posterior transpedicular screws and postoperative physiotherapy compared to cognitive intervention and exercises[22,23]; both publications report that "men had inferior results after surgery." However, these results were quantified neither in the surgery nor non-surgery groups.[22,23]

Our review also identified one good quality case series describing outcomes in a population of 106 patients with discogenic back pain treated with ALIF and followed for a mean of 29.7 months.[46] Using a multivariate regression model, the authors evaluated the impact of gender on a number of different outcome measures: returning to work, a 30% improvement in the VAS pain score, or an increase of at least 30% on the Roland Morris score.[46] Gender was not found associated with any of these outcomes.[46]

Outside the scope of this review, there are several sources of evidence which may add to our understanding of the relationship between gender and the outcomes of treatment for DDD. In a retrospective study[37] of 143 active duty military personnel (mean age 36.3 years), of whom 118 (83%) were DDD patients, treated with TLIF and followed for a mean of 34.9 months, neither gender, nor smoking status, nor levels of arthrodesis were significantly associated with medical separation (an inability to remain on active duty).

Workers' Compensation

Our review identified one good quality RCT, and one good quality case-series describing workers compensation as a potential predictor of the impact of DDD treatment. The RCT[29] is a secondary analysis of data derived from the Swedish Lumbar Spine Study[25,38]. The authors found that WC status was negatively associated with patient global assessment (p=0.049) and work status (p=0.035) in the surgical group, but not in the non-surgical group, indicating a differential impact of WC on treatment.[29]

Our review also identified one good quality case-series describing outcomes in a population of 106 patients with discogenic back pain treated with ALIF and followed for a mean of 29.7 months.[46] Using a multivariate regression model, the authors evaluated the impact of WC on a number of different outcome measures: returning to work, a 30% improvement in the VAS pain score, or an increase of at least 30% on the Roland Morris score (measuring function).[46] WC status was not found to be statistically-significantly associated with any of these outcomes. The multivariate model also included pre-surgery work status as a potential predictor of outcome, and this was independently associated with

return to work (OR 10.5 [2.64,41.4], p=0.0008), but not with VAS pain score or Roland Morris function score.[46]

Outside of the scope of this review are several sources of information which may further illustrate the variation in findings around the impact of WC status on the outcome of treatment of back pain patients. In contrast with the inconsistent findings above, compensation status, whether through litigation or workers' compensation, is in general consistently associated with poor outcomes after any surgical intervention, as reported in a systematic review of 211 clinical trials with relevant information.[70] Several relevant publications describe primary studies of lumbar fusion which add additional specific evidence to the association of WC and outcomes in groups treated thusly. A nonrandomized comparative prospective study of 66 patients undergoing single level TLIF compared those experiencing open (n=33, of which 14 were patients with DDD) vs minimally invasive surgery (n=33, of which 13 were patients with DDD).[54] This study found no significant differences in clinical outcomes between those receiving WC compared to the non-WC group, either overall, or stratified by the open versus minimally invasive technique.[54] These findings were in contrast to those of a prospective non-comparative study of 125 patients undergoing ALIF over a 2-year period (of whom 27 were patients with uncomplicated DDD), which documented a significantly lower rate of clinical success (as defined by a score of 1 or 2 on the PSI) among patients receiving WC (68% success rate) compared to those not (91% success rate) (p=0.006).[71] This negative relationship did not hold true in the analysis of either the ODI or the SF-12 Physical Component Summary (PCS) or Mental Component Summary (MCS).[71]

Psychological Factors

Our review identified two good quality studies describing psychosocial factors as potential predictors of the impact of treatment. The first was performed in the context of a good quality multicenter RCT.[29] This study[29] is a secondary analysis of data derived from the Swedish Lumbar Spine Study[25,38]. In this analysis of data from 294 enrolled patients, the authors evaluated factors they deemed as potential predictors of various treatment outcomes in surgical and conservative (non-intensive physical therapy) patient groups.[29] Outcome measures included reduction of disability (≥50% reduction of the ODI score), patient global assessment of treatment effect (improvement/no improvement), and work status at the conclusion of 2 years of followup.[29] Using a stepwise, forward multiple logistic regression analysis, the authors found that neurotic personality (measured using the Karolinska Scales of Personality) was statistically-significantly negatively associated with improvement in patient global assessment in the surgical group (p=0.006).[29] However, this association was not significant in the non-surgical group, indicating a differential impact of neurotic personality traits on treatment.[29]

Conversely, in this same study, depressive symptoms measured using the Zung Depression Scale were negatively associated with improvement in the patient global assessment score in the conservative group but not in the surgical group, suggesting as well a differential impact of this trait on treatment.[29] There was no association, differential or otherwise, noted between depression and either ODI or work status in either the surgical or non-surgical treatment groups.[29]

Our review also identified a good quality retrospective case series of 620 patients with DDD treated with single level posterolateral fusion, followed for at least 3 years, of whom 24.4% returned to work in within two years postoperatively.[35] Negative predictors of return to work included psychological comorbidity (defined as undergoing psychotherapy) before fusion (OR 0.30; 95% CI: 0.14, 0.62).[35]

Outside of the scope of the current review, there are data which may further illustrate nuances of the relationship between psychological comorbidities and outcomes of treatment for uncomplicated DDD.

A systematic literature review documented that psychological factors may in fact modify the treatment effect of fusion versus conservative treatment, with the outcome of fusion less favorable among patients with personality disorder, neuroticism, or depression.[72] Supporting these findings is a retrospective population-based cohort study[60] of 2,378 chronic back pain patients treated with lumbar fusion surgery and receiving Washington State workers compensation followed by for a mean of 6.6 years. This study[60] reports that psychological comorbidities, characterized as including depression, dysthymia, manic-depressive disorders, stress, affective psychoses, or adjustment disorders, were associated with a higher risk of disability two years after lumbar fusion (OR 1.51 [1.05-2.26]).

Lifestyle Factors

Smoking

Our review identified one good quality case series describing the impact of smoking on ALIF outcomes.[46] Smoking was not found to be associated with returning to work, a 30% improvement in the VAS pain score, or an increase of at least 30% on the Roland Morris score.

Outside the scope of this review, other publications also note no impact of smoking on outcomes of surgical treatment of back pain. In a retrospective study of 143 active duty military personnel (mean age 36.3 years), of whom 118 (83%) were DDD patients, treated with TLIF and followed for a mean of 34.9 months, smoking status was not significantly associated with medical separation (an inability to remain on active duty.[37]

BMI

Our review did not identify any studies presenting BMI-specific data as a characteristic of interest in patients with uncomplicated DDD. Outside the scope of this review, however, there is literature describing the association of BMI with the outcome of surgical treatment for back pain. A prospective non-comparative study of 125 patients undergoing ALIF over a 2-year period (of whom 27 were patients with uncomplicated DDD), documented no significantly different rates of clinical success (as defined by a score of 1 or 2 on the PSI) among patients in varying BMI strata, nor differences in the ODI and the PCS and MCS of the SF-12.[71]

Key Question #5: What are the costs and potential cost-effectiveness of lumbar fusion relative to alternative treatment approaches?

Economic evaluations of lumbar spinal fusion in patients with uncomplicated DDD are limited both in number and in quality. Available evidence on the costs of lumbar fusion surgery suggest that in-hospital costs alone can approach $100,000 in the U.S., particularly for more complex forms of surgery. The results of two RCT-based economic evaluations mirrored findings for clinical outcomes. A comparison of fusion to interdisciplinary rehabilitation in which no material differences in clinical effectiveness were observed yielded a two-year cost-effectiveness estimate of >$100,000 per quality-adjusted life-year gained. A second comparison of fusion to variable approaches for physical therapy produced calculated cost per unit improvement in pain and function as well as per case of symptom improvement or return to work rather than traditional cost-effectiveness measures such as unadjusted or quality-adjusted survival. Finally, a survey-based study of low back pain patients' willingness to pay for surgery indicated a willingness to pay more than the actual observed costs of surgery for discectomy and decompression alone, but not for lumbar fusion.

While many studies in the available literature have documented increases in both the utilization and costs of lumbar fusion surgery, relatively few have focused specifically on costs and potential cost-effectiveness in the target population for this assessment—patients with degenerative disc disease and chronic low back pain not attributable to other conditions (e.g., severe stenosis, acute trauma, etc.) and without radiculopathy. We summarize the available economic evidence for patients with uncomplicated DDD below, as well as those from selected other studies commenting on cost data and/or trends relevant to fusion surgery. Costs are presented in terms of 2014 US dollars, and were updated as necessary based on the medical care component of the U.S. Consumer Price Index.[73]

Utilization and Costs of Fusion in the U.S.

Given the policy interest around the use and appropriateness of fusion procedures in the U.S., it is not surprising that utilization of these procedures has been closely tracked. We chose to focus on comprehensive evaluations that have been performed most recently. One such study focused specifically on the use of lumbar fusion for DDD employed the Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample to evaluate trends from 2000-2009.[74] Population-adjusted utilization of fusion surgery increased 2.4-fold during this period, with the greatest increases seen in anterior approaches to fusion. Another relatively recent study used Medicare claims data to examine trends from 2002-2007 in utilization, outcome, and cost, although the focus of attention in this evaluation was on patients with spinal stenosis.[15] Results suggested a more than 15-fold increase (from 1.3 to 19.9 per 100,000 beneficiaries) in the rate of "complex" fusion procedures (i.e., more than two disk levels or a combined anterior/posterior approach), and an incidence of life-threatening complications with complex fusion (5.6%) more than 2-fold higher than among patients undergoing decompressive surgery without fusion. Adjusted hospital charges (in 2014 USD) ranged from $27,480 for decompression alone to $67,773 to complex fusion to $92,766 for complex fusion.

Martin and colleagues also explored whether differences in worker's compensation coverage policy for lumbar fusion in a variety of degenerative conditions had an impact on utilization and costs.[75] State inpatient databases were compared for California, which requires coverage in any situation in which a second opinion agrees with the first, and Washington, which applies utilization review criteria, requires imaging confirmation of spinal instability, and limits the initial procedure to a single disc level. In 2008-2009, the age- and sex-adjusted rate of lumbar fusion in the worker's compensation population was 19.0 per 100,000 employed adults in California and 12.9 per 100,000 in Washington ($p<0.001$). Rates of reoperation and readmission within three months of the initial procedure were also statistically-significantly higher in California. Finally, after adjustment for age, sex, comorbidity, and indication for fusion, mean hospitalization costs (2014 USD) were over 20% higher in California ($59,168 versus $48,271 for Washington, $p<0.001$).

Cost-Effectiveness of Lumbar Fusion in DDD

Two of the RCTs summarized in our assessment featured within-trial economic evaluations. In one, Rivero-Arias and colleagues evaluated the cost-effectiveness of lumbar fusion over a 2-year period[76] based on clinical, utility, and micro-costed data collected during Fairbank's RCT comparing lumbar fusion to intensive rehabilitation.[24] Costs were calculated based on itemized resources and unit costs for surgical, rehabilitation, and follow-up services utilized. Utility estimates were based on direct collection of data from the EuroQol EQ-5D questionnaire at multiple timepoints. Interestingly, while productivity loss was also costed, these estimates do not appear to have been used in the evaluation, which is described as having been conducted from the perspective of the British National Health Service.[76]

Two-year costs for surgery and rehabilitation (in 2014 US dollars) totaled $18,345 and $10,604 respectively. The difference in quality-adjusted survival between groups was 0.068 in favor of surgery (although this was not a statistically-significant difference). Cost-effectiveness (2014 USD) was $113,838 per quality-adjusted life year (QALY) gained for surgery. The authors concluded that such a ratio would not represent a cost-effective use of resources over a 2-year window, and sensitivity analyses suggested that cost-effectiveness might only be approached if differences in utility persisted over the long term and/or greater than 20% of rehabilitation patients opted for surgery each year.

The other trial-based evaluation comes from the Swedish Lumbar Spine Study[25] and also involved costing of resources consumed during the 2-year study.[77] Unfortunately, cost-effectiveness was expressed not in terms of cost per QALY or life-year gained, but in terms of unit improvements in disability, treatment success, and return to work. In primary analyses, cost-effectiveness of fusion (in 2014 USD) versus usual care was estimated to be $2,363 per unit improvement on the ODI. Original cost-effectiveness calculations appeared to treat differences in return to work and significant clinical improvement as whole numbers rather than proportions. When considered as proportions (i.e., differences in the probability of these outcomes), cost-effectiveness was $54,527 per significant clinical improvement and $81,011 per return to work respectively.

We identified two additional cost-effectiveness evaluations that made use of clinical data, although not from studies that were considered for our evidence review. Adogwa and colleagues examined the cost-effectiveness of TLIF in 45 patients with grade 1 spondylolisthesis[78], while Glassman et al. assessed the cost-effectiveness of PLIF among patients with DDD as well as other conditions (e.g., disc herniation).[79] In both studies, however, costs and QALYs at two years were compared to those before surgery in the same population rather than to a control group receiving a contemporaneous intervention. In Adogwa's study, cost-effectiveness was estimated to be $46,428 per QALY gained (2014 USD) at two years.[78] In the Glassman evaluation, the cost-effectiveness of fusion (2014 USD) was $34,565 per QALY gained when only direct health care costs were considered and $56,443 per QALY gained when costs of lost productivity were added. Again, these ratios are calculated in relation to a pre-surgical state rather than to the costs and outcomes associated with an alternative treatment.

Other Economic Evaluations

Fayssoux and colleagues estimated the indirect costs associated with surgery for single-level DDD by using pooled data from an RCT of lumbar fusion and artificial disc replacement.[80] In the first year postoperatively, rates of full- or part-time employment declined from approximately 54% at baseline to less than 30% at 6 weeks, but returned to baseline levels by one year. Lost wages totaled approximately $2,900 per patient in the first year. By the end of the second year of follow-up, 63% of patients reported full- or part-time employment.

Another study involved the use of a post-surgery evaluation of the value that patients ascribe to individual surgical procedures for low back pain.[81] A total of 115 Swiss patients who had undergone discectomy, decompressive surgery, or fusion for a variety of degenerative lumbar conditions were surveyed regarding the maximum they would be willing to pay for each of these procedures, controlling for other factors such as satisfaction with the procedure, family income, and other financial resources. For both discectomy and decompression, the maximum willingness-to-pay (WTP) threshold for surgery was higher than the actual cost of the surgical procedures. For lumbar fusion, however, patients reported a maximum willingness-to-pay level of $19,712 (2014 USD), compared to an actual average hospital cost of $24,676 ($p<0.05$).

Finally, Alvin and colleagues conducted a systematic review to document variation in cost-calculation methods in economic evaluations of cervical and spinal lumbar surgery.[82] A total of 37 economic evaluations were identified. Sources of costs varied widely, with approximately one-third of evaluations using public-payer reimbursement, another one-third based on procedure micro-costing approaches, and the remainder using cost-to-charge ratios or other government data sources. Of perhaps greater concern, one-quarter of the cost-effectiveness evaluations that stated use of a societal perspective did not include calculations of indirect costs, and there was great variation in the types of direct costs considered.

9. Recommendations for Future Research

Evidence reviewed in this assessment suggests that, overall, lumbar fusion does not provide incremental clinical benefit in comparison to various forms of coordinated and interdisciplinary rehabilitation programs in patients with uncomplicated degenerative disc disease. Even where benefits were seen (i.e., in comparison to less-intense forms of physical and exercise therapy), they tended to diminish over longer periods of follow-up.

While these findings seem to relegate the use of lumbar fusion to treatment of very last resort in patients with uncomplicated DDD, there are still unanswered questions regarding treatment alternatives in this patient population. For one, the literature comparing interdisciplinary rehabilitation to other forms of conservative management has produced inconsistent results. A 2011 ICER appraisal[92] attempted to identify the components of interdisciplinary rehabilitation most closely associated with treatment success. Not only were these components difficult to quantify, but available evidence suggests inconsistent effects for interdisciplinary rehabilitation on pain, function, and return to work when compared to usual care, and no material clinical benefits when compared to physical therapy alone. The field requires further refinement to define the characteristic components of interdisciplinary rehabilitation so that (a) programs can be compared on an equal footing; and (b) a minimum set of components can be identified for successful program application in community-based as well as more heavily-resourced settings. In addition, a measure of success reported too-infrequently in rehabilitation studies is avoidance of surgery itself. This should be a standard component of clinical trials moving forward.

There is also a dearth of evidence comparing lumbar fusion to minimally-invasive treatment alternatives in patients with uncomplicated DDD. While the evidence for some of these alternatives has also been questioned, stakeholders would benefit from understanding whether a “stepped care” approach would benefit patients with longstanding back pain.

Finally, the only way to understand whether fusion has any place in treatment is to conduct studies in more broadly-representative populations. This could be accomplished through either RCTs or registry-studies, and should involve both community and academic settings, multiple types of insurance coverage, and patients with back pain of varying duration. These studies could also be coupled with evaluations (randomized or otherwise) of screening protocols designed to triage patients toward certain types of treatments. For example, such protocols might appropriately steer patients with pronounced psychological distress or fear of activity toward the educational components of rehabilitation, while those with refractory pain despite exercise and physical therapy might derive greater benefit from more invasive treatment.

References

1. WHO. Priority medicines for Europe and the World Update Report, 2013. Priority diseases and reasons for inclusion: low back pain. 2013; http://www.who.int/medicines/areas/priority_medicines/Ch6_24LBP.pdf. Accessed March 2015, 2015.
2. Licciardone JC. The epidemiology and medical management of low back pain during ambulatory medical care visits in the United States. *Osteopath Med Prim Care.* 2008;2:11.
3. Soni A. Back Problems: Use and Expenditures for the U.S. Adult Population, 2007. Medical Expenditure Panel Survey. 2010. http://meps.ahrq.gov/mepsweb/data_files/publications/st289/stat289.pdf.
4. Stewart WF, Ricci JA, Chee E, Morganstein D. Lost productive work time costs from health conditions in the United States: results from the American Productivity Audit. *J Occup Environ Med.* 2003;45(12):1234-1246.
5. Aching backs impact cost, disability. *AAOS Now* 2010; http://www.aaos.org/news/aaosnow/jan09/research6.asp. Accessed March 2015, 2015.
6. Hestbaek L, Leboeuf-Yde C, Manniche C. Low back pain: what is the long-term course? A review of studies of general patient populations. *Eur Spine J.* 2003;12(2):149-165.
7. Johannes CB, Le TK, Zhou X, Johnston JA, Dworkin RH. The prevalence of chronic pain in United States adults: results of an Internet-based survey. *J Pain.* 2010;11(11):1230-1239.
8. Pengel LH, Herbert RD, Maher CG, Refshauge KM. Acute low back pain: systematic review of its prognosis. *BMJ.* 2003;327(7410):323.
9. Heliovaara M, Impivaara O, Sievers K, et al. Lumbar disc syndrome in Finland. *J Epidemiol Community Health.* 1987;41(3):251-258.
10. Ullrich P. Lumbar Degenerative Disc Disease (DDD). http://www.spine-health.com/conditions/degenerative-disc-disease/lumbar-degenerative-disc-disease-ddd. Accessed March 2015, 2015.
11. Eck JC, Sharan A, Resnick DK, et al. Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 6: discography for patient selection. *J Neurosurg Spine.* 2014;21(1):37-41.
12. Ragab A, Deshazo RD. Management of back pain in patients with previous back surgery. *Am J Med.* 2008;121(4):272-278.
13. Spinal Fusion. 2010; http://orthoinfo.aaos.org/topic.cfm?topic=a00348. Accessed March 2015, 2015.
14. Mayo Clinic. Laminectomy. http://www.mayoclinic.com/health/laminectomy/MY00674. Accessed August, 2015.
15. Deyo RA, Mirza SK, Martin BI, Kreuter W, Goodman DC, Jarvik JG. Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA.* 2010;303(13):1259-1265.
16. Schonstein E, Kenny D, Keating J, Koes B, Herbert RD. Physical conditioning programs for workers with back and neck pain: a cochrane systematic review. *Spine (Phila Pa 1976).* 2003;28(19):E391-395.

17. Chou R, Baisden J, Carragee EJ, Resnick DK, Shaffer WO, Loeser JD. Surgery for low back pain: A review of the evidence for an American pain society clinical practice guideline. *Spine.* 2009;34(10):1094-1109.
18. Sidhu GS, Henkelman E, Vaccaro AR, et al. Minimally invasive versus open posterior lumbar interbody fusion: a systematic review. *Clin Orthop Relat Res.* 2014;472(6):1792-1799.
19. Ollendorf DA, Pearson SD. An integrated evidence rating to frame comparative effectiveness assessments for decision makers. *Med Care.* 2010;48(6 Suppl):S145-152.
20. Ohtori S, Koshi T, Yamashita M, et al. Surgical versus nonsurgical treatment of selected patients with discogenic low back pain: A small-sized randomized trial. *Spine.* 2011;36(5):347-354.
21. ECRI. Spinal fusion and discography for chronic low back pain and uncomplicated lumbar degenerative disc disease. 2007; http://www.hca.wa.gov/hta/documents/spinal_fusion_discography_final_101907.pdf. Accessed March 2015, 2015.
22. Brox JI, Reikeras O, Nygaard O, et al. Lumbar instrumented fusion compared with cognitive intervention and exercises in patients with chronic back pain after previous surgery for disc herniation: a prospective randomized controlled study. *Pain.* 2006;122(1-2):145-155.
23. Brox JI, Sorensen R, Friis A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine.* 2003;28(17):1913-1921.
24. Fairbank J, Frost H, Wilson-MacDonald J, Yu LM, Barker K, Collins R. Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. *BMJ.* 2005;330(7502):1233.
25. Fritzell P, Hagg O, Wessberg P, Nordwall A. 2001 Volvo award winner in clinical studies: Lumbar fusion versus nonsurgical treatment for chronic low back pain. A multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. *Spine.* 2001;26(23):2521-2534.
26. Brox JI, Nygaard OP, Holm I, Keller A, Ingebrigtsen T, Reikeras O. Four-year follow-up of surgical versus non-surgical therapy for chronic low back pain. *Ann Rheum Dis.* 2010;69(9):1643-1648.
27. Froholdt A, Brox JI, Reikeras O, Leivseth G. Disc height and sagittal alignment in operated and non-operated levels in the lumbar spine at long-term follow-up: a case-control study. *Open Orthop J.* 2013;7:258-263.
28. Froholdt A, Reikeraas O, Holm I, Keller A, Brox JI. No difference in 9-year outcome in CLBP patients randomized to lumbar fusion versus cognitive intervention and exercises. *Eur Spine J.* 2012;21(12):2531-2538.
29. Hagg O, Fritzell P, Ekselius L, Nordwall A. Predictors of outcome in fusion surgery for chronic low back pain. A report from the Swedish Lumbar Spine Study. *Eur Spine J.* 2003;12(1):22-33.
30. Kang MS, Park JY, Kim KH, et al. Minimally invasive transforaminal lumbar interbody fusion with unilateral pedicle screw fixation: comparison between primary and revision surgery. *Biomed Res Int.* 2014;2014:919248.
31. Mannion AF, Brox JI, Fairbank JC. Comparison of spinal fusion and nonoperative treatment in patients with chronic low back pain: long-term follow-up of three randomized controlled trials. *Spine J.* 2013;13(11):1438-1448.

32. Mannion AF, Leivseth G, Brox JI, Fritzell P, Hagg O, Fairbank JC. ISSLS Prize winner: Long-term follow-up suggests spinal fusion is associated with increased adjacent segment disc degeneration but without influence on clinical outcome: results of a combined follow-up from 4 randomized controlled trials. *Spine.* 2014;39(17):1373-1383.
33. Mirza SK, Deyo RA, Heagerty PJ, Turner JA, Martin BI, Comstock BA. One-year outcomes of surgical versus nonsurgical treatments for discogenic back pain: a community-based prospective cohort study. *Spine J.* 2013;13(11):1421-1433.
34. Smith JS, Sidhu G, Bode K, et al. Operative and nonoperative treatment approaches for lumbar degenerative disc disease have similar long-term clinical outcomes among patients with positive discography. *World Neurosurg.* 2014;82(5):872-878.
35. Anderson JT, Haas AR, Percy R, Woods ST, Ahn UM, Ahn NU. Single-level lumbar fusion for degenerative disc disease is associated with worse outcomes compared with fusion for spondylolisthesis in a workers' compensation setting. *Spine (Phila Pa 1976).* 2015;40(5):323-331.
36. Lammli J, Whitaker MC, Moskowitz A, et al. Stand-alone anterior lumbar interbody fusion for degenerative disc disease of the lumbar spine: results with a 2-year follow-up. *Spine.* 2014;39(15):E894-901.
37. Schoenfeld AJ, Thomas D, Bader JO, Bono CM. Transforaminal lumbar interbody fusion: prognostic factors related to retention in an active duty military population. *Mil Med.* 2013;178(2):228-233.
38. Fritzell P, Hagg O, Wessberg P, Nordwall A. Chronic low back pain and fusion: A comparison of three surgical techniques: A prospective multicenter randomized study from the Swedish Lumbar Spine Study Group. *Spine.* 2002;27(11):1131-1141.
39. Washington State Health Care Authority. Spinal fusion and discography for chronic low back pain and uncomplicated lumbar degenerative disc disease. *ECRI Institute.* 2007.
40. Mirza SK, Deyo RA. Systematic review of randomized trials comparing lumbar fusion surgery to nonoperative care for treatment of chronic back pain. *Spine.* 2007;32(7):816-823.
41. Bydon M, De la Garza-Ramos R, Macki M, Baker A, Gokaslan AK, Bydon A. Lumbar fusion versus nonoperative management for treatment of discogenic low back pain: a systematic review and meta-analysis of randomized controlled trials. *J Spinal Disord Tech.* 2014;27(5):297-304.
42. Carreon LY, Glassman SD, Howard J. Fusion and nonsurgical treatment for symptomatic lumbar degenerative disease: a systematic review of Oswestry Disability Index and MOS Short Form-36 outcomes. *Spine J.* 2008;8(5):747-755.
43. Saltychev M, Eskola M, Laimi K. Lumbar fusion compared with conservative treatment in patients with chronic low back pain: a meta-analysis. *Int J Rehabil Res.* 2014;37(1):2-8.
44. Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976).* 2008;33(1):90-94.
45. Carragee EJ, Cheng I. Minimum acceptable outcomes after lumbar spinal fusion. *Spine J.* 2010;10(4):313-320.
46. Anderson PA, Schwaegler PE, Cizek D, Leverson G. Work status as a predictor of surgical outcome of discogenic low back pain. *Spine.* 2006;31(21):2510-2515.
47. Miyagi M, Ikeda O, Ohtori S, et al. Additional decompression at adjacent segments leads to adjacent segment degeneration after PLIF. *Eur Spine J.*22(8):1877-1883.

48. Fineberg SJ, Nandyala SV, Kurd MF, et al. Incidence and risk factors for postoperative ileus following anterior, posterior, and circumferential lumbar fusion. *Spine J.* 2014;14(8):1680-1685.
49. Goz V, Weinreb JH, Schwab F, Lafage V, Errico TJ. Comparison of complications, costs, and length of stay of three different lumbar interbody fusion techniques: an analysis of the Nationwide Inpatient Sample database. *Spine J.* 2014;14(9):2019-2027.
50. Martin BI, Mirza SK, Franklin GM, Lurie JD, MacKenzie TA, Deyo RA. Hospital and surgeon variation in complications and repeat surgery following incident lumbar fusion for common degenerative diagnoses. *Health Serv Res.* 2013;48(1):1-25.
51. Juratli SM, Mirza SK, Fulton-Kehoe D, Wickizer TM, Franklin GM. Mortality after lumbar fusion surgery. *Spine.* 2009;34(7):740-747.
52. Stromqvist B, Fritzell P, Hagg O, Jonsson B, Sanden B. Swespine: the Swedish spine register : the 2012 report. *Eur Spine J.* 2013;22(4):953-974.
53. Marquez-Lara A, Nandyala SV, Fineberg SJ, Singh K. Cerebral vascular accidents after lumbar spine fusion. *Spine (Phila Pa 1976).* 2014;39(8):673-677.
54. Pelton MA, Phillips FM, Singh K. A comparison of perioperative costs and outcomes in patients with and without workers' compensation claims treated with minimally invasive or open transforaminal lumbar interbody fusion. *Spine (Phila Pa 1976).* 2012;37(22):1914-1919.
55. Brodano GB, Martikos K, Lolli F, et al. Transforaminal lumbar interbody fusion in degenerative disc disease and spondylolisthesis grade I: Minimally invasive versus open surgery. *Journal of Spinal Disorders and Techniques.* 2013.
56. Robinson Y, Michaelsson K, Sanden B. Instrumentation in lumbar fusion improves back pain but not quality of life 2 years after surgery. *Acta Orthop.* 2013;84:7-11.
57. Korovessis P, Repantis T, Baikousis A, Iliopoulos P. Posterolateral versus circumferential instrumented fusion for monosegmental lumbar degenerative disc disease using an expandable cage. *European Journal of Orthopaedic Surgery and Traumatology.* 2012;22(8):639-645.
58. Wu Y, Tang H, Li Z, Zhang Q, Shi Z. Outcome of posterior lumbar interbody fusion versus posterolateral fusion in lumbar degenerative disease. *J Clin Neurosci.* 2011;18(6):780-783.
59. Liu Z, Fei Q, Wang B, et al. A meta-analysis of unilateral versus bilateral Pedicle screw fixation in minimally invasive lumbar Interbody fusion. *PLoS ONE.* 2014;9(11).
60. Maghout Juratli S, Franklin GM, Mirza SK, Wickizer TM, Fulton-Kehoe D. Lumbar fusion outcomes in Washington State workers' compensation. *Spine.* 2006;31(23):2715-2723.
61. Buttermann GR, Mullin WJ. Two-level circumferential lumbar fusion comparing midline and paraspinal posterior approach: 5-year interim outcomes of a randomized, blinded, prospective study. *Journal of Spinal Disorders and Techniques.* 2013.
62. Hoy K, Bunger C, Niederman B, et al. Transforaminal lumbar interbody fusion (TLIF) versus posterolateral instrumented fusion (PLF) in degenerative lumbar disorders: a randomized clinical trial with 2-year follow-up. *Eur Spine J.*22(9):2022-2029.
63. Faundez AA, Schwender JD, Safriel Y, et al. Clinical and radiological outcome of anterior-posterior fusion versus transforaminal lumbar interbody fusion for symptomatic disc degeneration: A retrospective comparative study of 133 patients. *Eur Spine J.* 2009;18(2):203-211.
64. Grob D, Bartanusz V, Jeszenszky D, et al. A prospective, cohort study comparing translaminar screw fixation with transforaminal lumbar interbody fusion and pedicle screw fixation for fusion

of the degenerative lumbar spine. *Journal of Bone and Joint Surgery - Series B.* 2009;91(10):1347-1353.

65. Dufour N, Thamsborg G, Oefeldt A, Lundsgaard C, Stender S. Treatment of chronic low back pain: a randomized, clinical trial comparing group-based multidisciplinary biopsychosocial rehabilitation and intensive individual therapist-assisted back muscle strengthening exercises. *Spine (Phila Pa 1976).* 2010;35(5):469-476.
66. Roche G, Ponthieux A, Parot-Shinkel E, et al. Comparison of a functional restoration program with active individual physical therapy for patients with chronic low back pain: a randomized controlled trial. *Arch Phys Med Rehabil.* 2007;88(10):1229-1235.
67. Guzman J, Esmail R, Karjalainen K, Malmivaara A, Irvin E, Bombardier C. Multidisciplinary rehabilitation for chronic low back pain: systematic review. *BMJ.* 2001;322(7301):1511-1516.
68. Ravenek MJ, Hughes ID, Ivanovich N, et al. A systematic review of multidisciplinary outcomes in the management of chronic low back pain. *Work.* 2010;35(3):349-367.
69. Anderson JT, Haas AR, Percy R, Woods ST, Ahn UM, Ahn NU. Clinical depression is a strong predictor of poor lumbar fusion outcomes among workers' compensation subjects. *Spine (Phila Pa 1976).* 2015;40(10):748-756.
70. Harris I, Mulford J, Solomon M, van Gelder JM, Young J. Association between compensation status and outcome after surgery: a meta-analysis. *JAMA.* 2005;293(13):1644-1652.
71. Rao PJ, Loganathan A, Yeung V, Mobbs RJ. Outcomes of anterior lumbar Interbody fusion surgery based on indication: A prospective study. *Neurosurgery.* 2015;76(1):7-23.
72. Daubs MD, Norvell DC, McGuire R, et al. Fusion versus nonoperative care for chronic low back pain: do psychological factors affect outcomes? *Spine (Phila Pa 1976).* 2011;36(21 Suppl):S96-109.
73. Consumer Price Index. 2015; http://www.bls.gov/cpi/tables.htm. Accessed August 2015.
74. Yoshihara H, Yoneoka D. National trends in the surgical treatment for lumbar degenerative disc disease: United States, 2000 to 2009. *Spine Journal.* 2015;15(2):265-271.
75. Martin BI, Franklin GM, Deyo RA, Wickizer TM, Lurie JD, Mirza SK. How do coverage policies influence practice patterns, safety, and cost of initial lumbar fusion surgery? A population-based comparison of workers' compensation systems. *Spine Journal.* 2014;14(7):1237-1246.
76. Rivero-Arias O, Campbell H, Gray A, Fairbank J, Frost H, Wilson-MacDonald J. Surgical stabilisation of the spine compared with a programme of intensive rehabilitation for the management of patients with chronic low back pain: cost utility analysis based on a randomised controlled trial. *BMJ.* 2005;330(7502):1239.
77. Fritzell P, Hagg O, Jonsson D, Nordwall A. Cost-effectiveness of lumbar fusion and nonsurgical treatment for chronic low back pain in the Swedish Lumbar Spine Study: a multicenter, randomized, controlled trial from the Swedish Lumbar Spine Study Group. *Spine (Phila Pa 1976).* 2004;29(4):421-434; discussion Z423.
78. Adogwa O, Parker SL, Davis BJ, et al. Cost-effectiveness of transforaminal lumbar interbody fusion for Grade I degenerative spondylolisthesis.[Erratum appears in J Neurosurg Spine. 2011 Aug;15(2):211; PMID: 21682560]. *J Neurosurg Spine.*15(2):138-143.
79. Glassman S, Gornet MF, Branch C, et al. MOS Short Form 36 and Oswestry Disability Index outcomes in lumbar fusion: A multicenter experience. *Spine Journal.* 2006;6(1):21-26.

80. Fayssoux R, Goldfarb NI, Vaccaro AR, Harrop J. Indirect costs associated with surgery for low back pain-a secondary analysis of clinical trial data. *Popul Health Manag.*13(1):9-13.
81. Haefeli M, Elfering A, McIntosh E, Gray A, Sukthankar A, Boos N. A cost-benefit analysis using contingent valuation techniques: A feasibility study in spinal surgery. *Value in Health.* 2008;11(4):575-588.
82. Alvin MD, Miller JA, Lubelski D, et al. Variations in cost calculations in spine surgery cost-effectiveness research. *Neurosurg.*36(6):E1.
83. Christensen FB. Lumbar spinal fusion. Outcome in relation to surgical methods, choice of implant and postoperative rehabilitation. *Acta Orthop Scand Suppl.* 2004;75(313):2-43.
84. Herkowitz HN, Sidhu KS. Lumbar Spine Fusion in the Treatment of Degenerative Conditions: Current Indications and Recommendations. *J Am Acad Orthop Surg.* 1995;3(3):123-135.
85. American Academy of Orthopaedic Surgeons. Spinal fusion. http://orthoinfo.aaos.org/topic.cfm?topic=a00348. Accessed August, 2015.
86. Phillips FM, Slosar PJ, Youssef JA, Andersson G, Papatheofanis F. Lumbar spine fusion for chronic low back pain due to degenerative disc disease: a systematic review. *Spine.* 2013;38(7):E409-422.
87. Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976).* 2000;25(22):2940-2952; discussion 2952.
88. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine (Phila Pa 1976).* 1983;8(2):141-144.
89. Chou R, Huffman LH, American Pain S, American College of P. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med.* 2007;147(7):492-504.
90. Wu RH, Fraser JF, Hartl R. Minimal access versus open transforaminal lumbar interbody fusion: Meta-analysis of fusion rates. *Spine.* 2010;35(26):2273-2281.
91. Liu J, Tang J, Liu H. Comparison of one versus two cages in lumbar interbody fusion for degenerative lumbar spinal disease: a meta-analysis. *Orthop Surg.* 2014;6(3):236-243.
92. ICER. *Management Options for Low Back Disorders.* Boston: Institute for Clinical and Economic Review; 2011.

Appendix A: Literature Search Strategy

Databases: MEDLINE, EMBASE, Cochrane Register of Controlled Trials, Databases of Abstracts of Reviews of Effects (DARE), OT Seeker, PEDro, ABI Inform, EconLit, and Health and Psychosocial Instruments
Search Date: June 1, 2015

Ovid Search Terms:
1. exp low back pain/ or exp lumbar vertebrae/ or exp lumbosacral region/
2. (low back pain or lumbar or lumbar spine or lumbosacral).ti,ab
3. 1 or 2
4. exp arthrodesis/
5. (arthrodesis or lumbar fusion* or interbody or posterolateral).ti,ab
6. 4 or 5
7. exp intervertebral disc degeneration/
8. (dis* degenerat* or degenerat* dis* or discogenic).ti,ab.
9. 7 or 8
10. 3 and 6 and 9
11. limit to (humans and english language and yr="2000 -Current")

Embase Search Terms:
1. ‘'lumbar vertebra'/exp OR 'lumbar':ab,ti OR 'lumbar spine':ab,ti OR 'lumbosacral':ab,ti OR 'low back pain'/exp OR 'low back pain':ab,ti
2. 'arthrodesis'/de OR 'spine fusion'/exp OR 'arthrodesis':ab,ti OR 'lumbar fusion':ab,ti OR 'interbody':ab,ti OR 'posterolateral':ab,ti
3. 'intervertebral disk degeneration'/de OR 'intervertebral disk degeneration'/de OR (dis* NEXT/1 degenerat*):ab,ti OR (degenerat* NEXT/1 dis*):ab,ti OR 'discogenic pain'/de OR 'discogenic':ab,ti
4. #1 AND #2 AND #3
5. #4 AND [humans]/lim AND [english]/lim AND [2000-2015]/py

Include:

- **Population**: adults with chronic (≥3 months) lumbar pain and degenerative disk disease (also called spondylosis)
    - *Note: patients with a history of prior back surgery for any indication should also be included*
- **Interventions/Comparator**: all major technical approaches to lumbar fusion surgery, regardless of surgical technique or hardware utilized, versus nonsurgical management, including conservative approaches of varying intensity (e.g., physical therapy, intensive exercise/rehabilitation, cognitive behavioral therapy, medication) OR minimally-invasive treatments (e.g., radiofrequency ablation, electrothermal therapy)
- **Outcomes**: patient- and clinician-reported measures of pain, function, and disability; opioid medication use; requirements for repeat surgery/retreatment; return to work and/or resumption of normal activities; complications and mortality; costs
- **Sources**: systematic reviews & meta-analyses, RCTs, comparative cohort studies, case series with at least 100 patients and at least 2 years of follow-up

Exclude:

- **Population**: patients with other spinal conditions such as radiculopathy, >Grade 1 degenerative spondylolisthesis, isthmic spondylolisthesis, or severe spinal stenosis, as well as acute trauma or systemic disease affecting the lumbar spine
    - *Note: studies with mixed patient populations should be included ONLY if outcomes are reported separately for individuals with DDD alone*
- **Interventions/Comparator**: other surgical procedures, including discectomy/laminectomy and artificial disc replacement
    - *Note: keep if used in combination with lumbar fusion*
- **Outcomes**: general surgical outcomes (e.g., blood loss, response to anesthesia, operating time)
- **Sources**: case series with less than 100 patients and less than 2 years of follow-up; case reports; conference abstracts; letters; reviews (not systematic); dissertations

Appendix B: Summary Evidence Tables

Author	Study Design	Study Quality	Intervention	# of Patients	Inclusion Criteria	Patient Characteristics	Follow-up	Outcomes	Harms
Anderson 2006[46]	Case series	Poor	ALIF with titanium cages and autogenous iliac bone graft	n=106	CLBP ≥ 6 months	VAS: 6.8 (patients working) VAS: 7.2 (patients not working) p=NR RDQ: 12.4 (patients working) RDQ: 16.5 (patients not working) p=NR	Mean 29.7 months with 95% follow-up after one year and 81% @ 24 months	VAS: -2.2 RDQ: -4.7 Workers' comp (# of patients working): preop yes - 13/50 no - 36/50 postop yes - 28/50 no - 22/50 Proportion patients working at follow-up who were working before surgery was 92% compared with 43% who did not work before surgery (p=0.0001, OR 10.5) and was independent of workers' comp status	Not reported
Brox 2003[23]	RCT	Good	1) fusion with posterior transpedicular screws and postoperative physiotherapy	n=64 1) 37 2) 27	Age 25–60 years Pain duration ≥1 year ≥30 on ODI Degeneration at L4–L5 and/or L5–S1	Age: 1) 44.1, 2) 42.4 % male: 1) 43, 2) 33 % smoking: 1) 41, 2) 44	12 months with 97% follow-up	ODI: 1) -15.6 2) -13.3 Difference of 2.3; after controlling for gender and pretreatment beliefs differences,	Early complications: 6/33 patients (18%) No late complications Mortality not reported

Author	Study Design	Study Quality	Intervention	# of Patients	Inclusion Criteria	Patient Characteristics	Follow-up	Outcomes	Harms
			2) cognitive intervention and exercises			Pain duration (years): 1) 9.5, 2) 12.5 ODI: 1) 41, 2) 42 Comorbidity (%): 1) 32, 2) 22		difference of 2.7, p=NS for both GFS score (0-100): 1) -17.6 2) -22 Back pain: 1) -22.7 2) -15.4 % working: 1) +2 2) +11	
Brox 2006[22]	RCT	Good	1) fusion with posterior transpedicular screws 2) cognitive intervention and exercises	n=60 1) 29 2) 31	Age 25–60 years Pain duration ≥ 1 year Disc degeneration at L4–L5 and/or L5–S1 All patients had undergone prior surgery for disc herniation	Age: 1) 42, 2) 43 % male: 1) 38, 2) 65 % smoking: 1) 72, 2) 58 ODI: 1) 47.0 2) 45.1	12 months	ODI: 1) -8.9 2) -12.8 Difference of -3.7, 95% CI, -13.5 to 6.2 After controlling for gender and pretreatment beliefs differences, difference of -7.3, 95% CI: -21.7, 1.7) ODI scores for the men in the surgery group did not improve -19.8 (-32.8 to -6.8), p=0.001 The 2 men who had surgery after follow-up did not improve	Early complications: 2/23 (8.7%) (due to wound infections) No late complications One death during follow-up (reason/group not reported)

Author	Study Design	Study Quality	Intervention	# of Patients	Inclusion Criteria	Patient Characteristics	Follow-up	Outcomes	Harms
								Patients randomized to cognitive intervention and exercise improved significantly from baseline to 1-year follow-up in all variables except back pain (p=0.07), RTW (p=0.13), and emotional distress (p=0.08)	
Brox 2010[26]	Follow-up to Brox 2003 and 2006	Good	1) fusion with posterior transpedicular screws 2) cognitive intervention and exercises	n=124 1) 66 2) 58 (merged patients from prior randomized trials)	Age 25-60 years Pain duration ≥1 year Disc degeneration at L4–L5 and/or L5–S1 ≥1 year of symptoms after or without previous surgery for disc herniation	Age: 1) 42.7, 2) 42.4 % male: 1) 41, 2) 50 % previous surgery: 1) 44, 2) 53 % smoking: 1) 36, 2) 30 ODI: 1) 44.4, 2) 43.0	48 months 1) 92% follow-up 2) 86% follow-up p=NR	ITT ODI: 1) -14.4, 2) -16.4 Adjusted treatment effect of 1.1; 95% CI: 5.9, 8.2 As-treated ODI: 1) -15.3 2) -15.3 Adjusted treatment effect -1.6; 95% CI: −8.9, 5.6 (outcomes adjusted for age, gender, baseline score and previous disc surgery) RTW: More patients	Reoperations: 1) 4 (25%) (crossover patients) 2) 15 (25%) 3 (5%) patients died in fusion group (unrelated to surgery) No other complications reported

Author	Study Design	Study Quality	Intervention	# of Patients	Inclusion Criteria	Patient Characteristics	Follow-up	Outcomes	Harms
								who had surgery (53% vs. 32%) were on disability pension (adjusted OR 2.5; 95% CI: 1.1, 5.9)	
Fairbank 2005[24]	RCT	Good	1) fusion (choice of technique was allowed) 2) intensive rehabilitation program (exercise targeted to individual's need)	n=349 1) 176 2) 173	Age 18-55 Pain duration at ≥1 year	Age: means reported for age ranges % male: 1) 44.9, 2) 53.8 % smoking: 1) 43.2, 2) 42.8 % previous surgery: 1) 8, 2) 8.1 ODI: 1) 46.5, 2) 44.8 SF-36 PCS: 1) 19.4, 2) 20 SF-36 MCS: 1) 43.2, 2) 44.8	24 months with 80% follow-up	ODI scores improved slightly more in favor of surgery (-4.1, 95% CI: -8.1, -0.1, p=0.045) After imputation for missing follow-up data the mean difference was -4.5 (95% CI: -8.2, 0.8, p=0.02) No differences between groups for surgery vs. control on any other outcome, including SF-36 PCS (+9.4 vs. +7.6), SF-36 MCS, (+4.2 vs. +0.7), or mental health (0-100 scale; +6.4 vs. +8.1)	Complications from surgery 19/176 (10.8%) Reoperations 11/176 (6.3%) Mortality not reported No complications associated with nonsurgical group
Fritzell 2001[25]	RCT	Fair	1) PLF, PLF with variable screw placement, ALIF or PLIF 2) nonsurgical (physical therapy,	n=294 1) 222 2) 72	Aged 25–65 years with CLBP Referred by PCP Pain ≥2 years from L4–L5	Age: 1) 43; 2) 44 % male: 1) 49.5; 2) 48.5 % smoking:	24 months with 98% follow-up Cross-over @ year 2 (n): 1) 18	Differences between groups for primary outcomes: VAS back: 16.7 (p=0.0002) VAS leg: 13.3 (p=0.005)	Early complications: 1) 53 (17%) Late complications: 13 (6%) Reoperations: 16

Author	Study Design	Study Quality	Intervention	# of Patients	Inclusion Criteria	Patient Characteristics	Follow-up	Outcomes	Harms
			cognitive training, pain relief)		and/or L5–S1	1) 40.6, 49.3 % previous surgery: 1) 18.6; 2) 19.4 1) VAS back: 64.2 VAS leg: 35.3 ODI: 47.3 Zung: 39.1 2) VAS back: 62.6 VAS leg: 35.6 ODI: 48.4 Zung: 39.4	2) 7	ODI: 8.8 (p=0.015) ZDS: 39.1 (p=NS) VAS & ODI scores in favor of surgery Compensation rating of "much better or "better" on litigation/compensation: 1) yes: 58%; no: 70% (p=NS) 2) yes: 18%; no: 58% (p=0.043)	(7.8%) unintended reoperations related to compilations in the fusion cohort 2 patients in the surgical group died within 2 years from baseline (not related to surgery)
Fritzell 2002[38]	Secondary analysis to Fritzell 2001	Fair	1a) PLF 1b) PLF w/VSP 1c) ALIF or PLIF	N=222 1a) 73 1b) 74 1c) 75	Aged 25–65 years with CLBP Referred by PCP Pain ≥2 years from L4–L5 and/or L5–S1 All randomized patients in surgical cohort from Fritzell 2001	Age: 43 % male: 49.5 % previous surgery: 18.6 VAS back: 64.2 VAS leg: 35.3 ODI: 47.3 ZDS: 39.1	24 months with 98% follow-up Cross-over @ year 2: 1a) 5 1b) 9 1c) 4	p=NS any outcome by surgery type, including ODI (-10.8, -14.8, -8.8), GFS (-12.3, -17.6, -16.3), ZDS (-8.8, -7.6, -7.1) for PLF, PLF with VSP, and ALIF/PLIF, respectively Work status-significant difference in favor of surgery expressed as “net back to work” (p=0.002) and also as “back to work” (p=0.049)	Early complications: overall: 17 1a) 6 1b) 16 1c) 31 1a vs. 1c, p<0.0001 Late complications (>2 weeks after surgery) 2 patients (implant-related infections) Reoperations: 1a) 2 1b) 0 2c) 1

Author	Study Design	Study Quality	Intervention	# of Patients	Inclusion Criteria	Patient Characteristics	Follow-up	Outcomes	Harms
								% net back to work by surgical type: 1a) 35 1b) 35 1c) 37	No complications in nonsurgical group
Froholdt 2012[28]	Follow-up to Brox 2010	Good	see Brox 2010	n=99 1) 55 2) 44	See Brox 2010	Age: 1) 43.0, 2) 42.6 % male: 1) 35, 2) 43 ODI: 1) 62.4, 2) 63.2 Pain duration: 1) 62.4, 2) 63.2 % smoking: 1) 55, 2) 51	9 years 80% from original study (Brox 2010) Patients who did not attend the 9-year follow-up were not different from dropouts at baseline or the latest follow-up before 9 years	RTW (full-time) 1) 35% 2) 36% p=NS ODI change (ITT analysis): 1) 20.2 2) 19.8 Adjusted treatment effect 1.9, 95% CI: 7.8, 11.6	Overall reoperations: 19/60 (32%) No infections and mortality not reported

Author	Study Design	Study Quality	Intervention	# of Patients	Inclusion Criteria	Patient Characteristics	Follow-up	Outcomes	Harms
Froholdt 2013[27]	Follow-up to Brox 2010	Good	see Brox 2010	n=48 1) 23 2) 25	See Brox 2010 Patients with baseline and follow-up x-rays	Age: 1) 44.0, 2) 41.7 % male: 1) 48, 2) 56 Pain duration (years): 1) 9.6, 2) 8.3 Back pain: 1) 65, 2) 60 % smoking: 1) 48 2) 48	9 years No significant differences in baseline characteristics between the patients included in the present study and all patients included in the original clinical trial	Not reported	p=NS for adjacent disc degeneration at 9 years follow-up; poor correlation between radiological adjacent segment degeneration and clinical symptoms ranging from r=0.04 (p=0.79) to r=0.36 (p=0.01)
Haag 2003[29]	Secondary analysis of Fritzell 2001	Good	1) fusion 2) nonsurgical (See Fritzell 2001 for details)	1) 201 2) 63	See Fritzell 2001	See Fritzell 2001	2 years	% with worker's comp who consider back problem better/not better after surgery, mean ODI for those who consider back problem better/not better after surgery, mean GFS for those who consider back problem better/not better were not different between groups % with worker's compensation who are working/not	Not reported

Author	Study Design	Study Quality	Intervention	# of Patients	Inclusion Criteria	Patient Characteristics	Follow-up	Outcomes	Harms
								working 1) 29/45; p=0.035 2) 39/19 Mean ZDS for back problem better/not better after surgery: 1) 39/39 2) 48/40; p=0.007 p=NS within groups unless reported; other baseline variables not statistically associated with improvement in back problem or work status	
Kang 2014[30]	Prospective cohort	Poor	Minimally invasive TLIF	n=46 1) 25 (primary) 2) 21 (revision)	DDD without instability including spinal stenosis, either primary or revision (prior discectomy)	Age: 1) 57.4, 2) 51.5 % male 1) 56, VAS leg pain 1) 7.9 2) 7.6 VAS back pain 1) 7.7 2) 7.7	1) 17.6 2) 16.3 p=0.45	VAS for leg pain: 1) -6.1 2) -6.2 VAS for back pain: 1) -5.7 2) -4.9 ODI: 1) -16.6 2) -14.8 (all outcomes measured at 1 year and p=NS)	Reoperation 1) 1 2) 0 Overall complications: 1) 3 (12%) 2) 4 (19%) Dural tear 1) 1 (4%) 2) 4 (19%) Cage migration 1) 2 (8%) 2) 0 (0%)

Author	Study Design	Study Quality	Intervention	# of Patients	Inclusion Criteria	Patient Characteristics	Follow-up	Outcomes	Harms
						ODI 1) 54.6 2) 66			Mortality not reported
Lammli 2014[36]	Case series	Poor	1-level or 2-level ALIF with recombinant human bone morphogenetic protein 2	n=118	Age 18-70 Failed conservative care >3 months; No psychological contraindications for surgery Completed 2-yr follow-up	Age: 43 % male: 41.5 % smoking: 33.1 % prior surgery: 32.2 VAS: 6.35	2 years	Average ODI improvement: 17%, p=0.036 VAS: -3.33, p<0.0001 Improved function, n (%): 60 (62.5) Maintained functionality, n (%): 25 (26.0) Worse functional outcomes n (%) : 11 (11.4)	No intraoperative or major complications Hematoma (n): 2 Additional surgical procedures (n): 9 Operations not related to adjacent level or fusion site (n): 3 Degeneration at level adjacent to fusion (n): 3 Pseudoarthrosis at fusion level (n): 3
Mannion 2013[31]	Follow-up to Brox 2003, Brox 2006, and Fairbank 2005	Good	1) fusion 2) nonsurgical (multidisciplinary cognitive-behavioral and exercise rehabilitation)	n=261 1) 140 2) 121	Participated in Brox 2003, Brox 2006, or Fairbank 2005 and consented to radiographical long-term follow-up	Age: 41.8 % male: 47.5 % previous spine surgery: 19.1 % smoking: 39.0	11.4 years	ITT ODI: 1) -14.8 2) -12.6 VAS back: 1) 44.3 2) 44.2 % taking everyday pain medication: 1) 22	Repeat surgery: 15% Hardware removal (n): 17 Further fusion (n): 11 Both hardware removal and repeat fusion (n): 9

Author	Study Design	Study Quality	Intervention	# of Patients	Inclusion Criteria	Patient Characteristics	Follow-up	Outcomes	Harms
						ODI: 1) 45.1 2) 42.3		2) 18 All outcome at long-term follow-up, p=NS As-treated outcomes also reported	 Deaths: 1) 10 2) 1 p=NR
Mannion 2014[32]	Follow-up to Fritzell 2001, Brox 2003, Brox 2006, and Fairbank 2005	Fair	1) fusion 2) nonsurgical (see Fritzell 2001, Brox 2003, Brox 2006, or Fairbank 2005)	n=369 1) 272 2) 97	Participated in Fritzell 2001, Brox 2003, Brox 2006, or Fairbank 2005 and consented to radiographical long-term follow-up	Age : 1) 43.1 2) 41.4 % male 1) 45.6 2) 49.5 % smoking: 1) 39.4 2) 42.7 ODI: 1) 46.7 2) 41.9 p=0.001	13.1 years	Not reported	Fusion associated with lower adjusted disc space height of the adjacent (cranial) segment of −0.44 SDs (95%CI −0.77 to −0.11; p=0.01) and of the next level up of −0.38 SDs (95% CI, −0.73 to −0.04; p=0.03); no influence on superior level No significant effects of adjacent level disc space height or posteroanterior displacement on VAS
Mirza 2013[33]	Prospective cohort	Good	1) surgery (instrumented fusion, artificial disc replacement, laminectomy or	n=495 1) 86* 2) 409 *68 (79%) fusion, 10	LBP for ≥6 months and MRI scan showing disc degeneration at one or two	Age: 1) 42.1, 2) 42.7 % male: 1) 45, 2) 48	12 months	30% improvement in pain intensity: 1) 71% 2) 35% p<0.001	Reoperation rate, n (%): 8 (11) No other complications or mortality reported

Author	Study Design	Study Quality	Intervention	# of Patients	Inclusion Criteria	Patient Characteristics	Follow-up	Outcomes	Harms
			discectomy) 2) nonsurgical (all those who did not receive surgery within 6 months	artificial disc replacement, 8 laminectomy or discectomy	lumbar discs	% previous surgery: 1) 36, 2) 21 p=0.004 % smoking: 1) 21, 2) 29		30% improvement in Roland score: 1) 57 2) 25 p<0.001 RTW (full- or part-time): 1) 59 2) 57 p=NS No opioid pain medications in past 3 months @ 1 year: 1) 47 2) 51 p=NS	
Ohtori 2011[20]	RCT	Fair	1) exercise treatment 2) anterior discectomy + ABF 3) PLF without decompression	n=41 1) 20 2) 15 3) 6	LBP ≥2 years with no accompanying radicular pain	Age: 1) 33, 2) 35, 3) 37 % male: 1) 50, 2) 66.7, 3) 66.7 Pain duration (years): 1) 7, 2) 7, 3) 9 VAS: 1) 7.7, 2) 7.4, 3) 6.5	1) 3 years 2) 4 years 3) 4 years p=NS	Change VAS after 2 years 1) -3.0 2) -6.1 3) -4 Change JOAS after 2 years 1) +0.5 2) +1.4 3) +1.3 Change ODI after 2 years 1) -24 2) -51.7 3) -44.8	Not reported

Author	Study Design	Study Quality	Intervention	# of Patients	Inclusion Criteria	Patient Characteristics	Follow-up	Outcomes	Harms
						JOA: 1) 0.7, 2) 1.1, 3) 0.7 ODI: 1) 64, 2) 62, 3) 66		(2) & (3) vs. (1) for all 3 outcomes: p<0.01 (2) vs. (3) for VAS and ODI: p<0.05	
Schoenfeld 2013[37]	Case series	Poor	TLIF	n=143	All TLIF procedures performed on active duty military personnel within Dept. of Orthopaedic Surgery Documented follow-up ≥1 year and/or Medical Evaluation Board (MEB) documentation	Age: 36.3 % male: 87 % Army: 95 % DDD: 83	34.9 months (42.5 months for those who were not medically separated)	Able to remain on active duty: 65% Younger individuals at increased risk of separation after TLIF OR 0.93 per each year increase in age (95% CI: 0.87, 0.98) Junior Enlisted personnel at increased risk of medical separation vs. Senior Enlisted and Officers: OR 6.42 (95% CI: 2.20, 18.74)	Sustained a complication, n (5): 7 (5%) Postoperative infection, n (%): 3 (2%) Seroma, n (5): 3 (2%) L5 radiculitis, n (%): (0.7%) Required revision: 7 (5%) Underwent MEB for medically separation, n (%): 50 (35%) Pseudoarthrosis, n (%): 6 (4%) Arthrodesis, n (%): 17 (12%)

Author	Study Design	Study Quality	Intervention	# of Patients	Inclusion Criteria	Patient Characteristics	Follow-up	Outcomes	Harms
									Fusion could not be reliably assessed (n): 54
Smith 2014[34]	Retrospective cohort	Poor	1) Fusion 2) Nonsurgical (physical therapy, epidural injections, medication)	n=96 1) 53 2) 43	Symptoms of axial low back pain, attempted conservative therapy for a minimum of 6 weeks, and a 1-level or a 2-adjacent level positive discogram that was concordant with lumbar DDD based on MRI	Age: 1) 47.0, 2) 47.3 NRS pain score: 1) 7.8, 2) 8.0 % male: 1) 47.2 , 2) 53.5	1) 63 months 2) 58 months p=NS	All patient-reported outcomes NRS pain score: 1) 3.6 2) 4.2 p=NS No preoperative ODI or SF-12 scores available but no differences at end of evaluation Smoking demonstrated a significant negative impact on ODI, SF-12 MCS, SF-12 PCS, and the satisfaction scale, and higher BMI had a significant negative impact on the satisfaction scale score.	Not reported

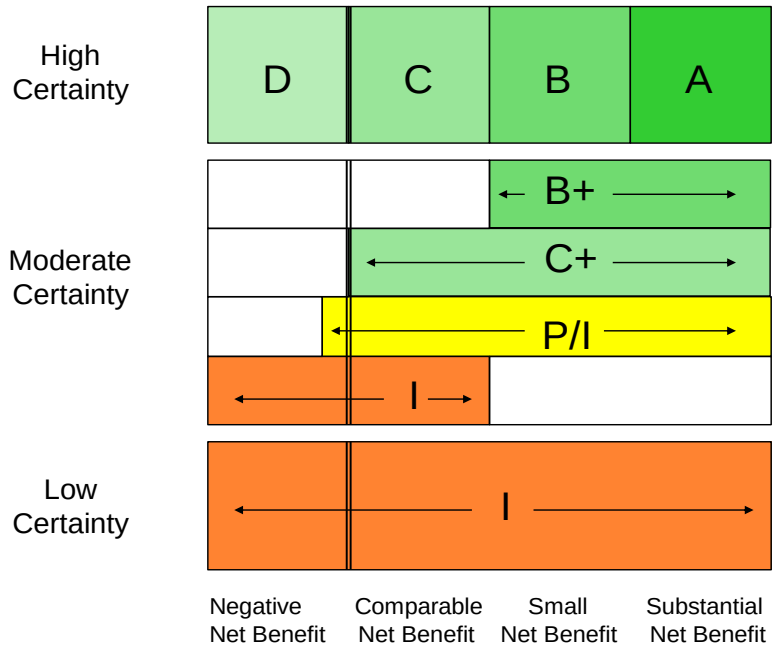
Appendix C: ICER Integrated Evidence Ratings

Formulary decisions require a rigorous evaluation of available evidence, a process that entails judgments regarding the quality of individual clinical studies and, ultimately, an assessment of the entire body of evidence regarding a therapeutic agent. To support this latter step, the Institute for Clinical and Economic Review (ICER) has developed the ICER Evidence Rating Matrix™. This user's guide to the ICER Matrix was developed with funding provided by the Comparative Effectiveness Research Collaborative Initiative (CER-CI), a joint initiative of the Academy of Managed Care Pharmacy, the International Society of Pharmacoeconomics and Outcomes Research, and the National Pharmaceutical Council (http://www.npcnow.org/issue/cer-collaborative-initiative). The ICER Matrix presents a framework for evaluating the comparative benefits and risks of therapies in a consistent, transparent system leading to an evidence rating that can guide coverage and formulary placement decisions. The purpose of this user's guide is to help members of Pharmacy and Therapeutics Committees and other decision-makers understand the approach embodied in the matrix, and to help them apply it in a reliable, consistent fashion.

The updated ICER Evidence Rating Matrix is shown below, with a key to the single letter ratings on the following page. Fundamentally, the evidence rating reflects a joint judgment of two critical components:

a) The **magnitude** of the difference between a therapeutic agent and its comparator in "net health benefit" – the balance between clinical benefits and risks and/or adverse effects (horizontal axis); AND

b) The level of **certainty** that you have in your best point estimate of net health benefit (vertical axis).



The letter ratings are listed below, according to the level of certainty in the best estimate of net health benefit.

High Certainty

A = Superior
B = Incremental
C = Comparable
D = Inferior

Moderate Certainty

B+=Incremental or Better
C+=Comparable or Better
P/I = Promising but Inconclusive
I = Insufficient

Low Certainty

I = Insufficient

Steps in Applying the ICER Evidence Rating Matrix

1. **Establish the specific focus of the comparison to be made and the scope of evidence you will be considering.** This process is sometimes referred to as determining the "PICO" – the **P**opulation, **I**ntervention, **C**omparator(s), and **O**utcomes of interest. Depending on the comparison, it is often helpful to also define the specific **T**ime Horizon and **S**etting that will be considered relevant.

2. **Estimate the magnitude of the comparative net health benefit.** Working from the scope of evidence established, it is important to quantify findings from the body of evidence on specific clinical benefits, risks, and other potentially important outcomes, such as adherence, so you can compare these side-by-side for the therapeutic agent and comparator. Some organizations compare each outcome, risk, etc. separately without using a quantitative measure to try to sum the overall comparative balance of benefits and risks between the therapeutic agent and the comparator. For these organizations the estimate of comparative net health benefit must be made qualitatively. Other organizations summarize the balance of benefits and risks using formal mathematical approaches such as health utility analysis, which generates a quantitative summary measure known as the quality-adjusted life year (QALY). What is most important, however, is full and transparent documentation of your rationale for assigning the magnitude of comparative net health benefit into one of four possible categories:

    - ***Negative:*** the drug produces a net health benefit inferior to that of the comparator
    - ***Comparable:*** the drug produces a net health benefit comparable to that of the comparator
    - ***Small:*** the drug produces a small positive net health benefit relative to the comparator
    - ***Substantial:*** the drug produces a substantial (moderate-large) positive net health benefit relative to the comparator

3. **Assign a level of certainty to the estimate of comparative net health benefit.** Given the strength of the evidence on comparative benefits and risks, a "conceptual confidence interval" around the original estimate of comparative net health benefit can be made, leading you to an assignment of the overall level of certainty in that estimate. Rather than assigning certainty by using a fixed equation weighting different attributes of the body of evidence, we recommend formal documentation of the consideration of 5 major domains related to strength of evidence: (1) Level of Bias—how much risk of bias is there in the study designs that comprise the entire evidence base? (2) Applicability—how generalizable are the results to real-world populations and conditions? (3) Consistency—do the studies produce similar treatment effects, or do they conflict in some ways? (4) Directness—are direct or indirect comparisons of therapies available, and/or are direct patient outcomes measured or only surrogate outcomes, and if surrogate outcomes only, how validated are these measures? (5) Precision—does the overall database include enough robust data to provide precise estimates of benefits and harms, or are estimates/confidence intervals quite broad?

   If you believe that your "conceptual confidence interval" around the point estimate of comparative net health benefit is limited to the boundaries of one of the four categories of comparative net health benefit above, your level of certainty is "high". "Moderate" certainty reflects conceptual confidence interval s extending across two or three categories, and may include drugs for which your conceptual confidence interval includes a small likelihood of a negative comparative net health benefit. When the evidence cannot provide enough certainty to limit your conceptual confidence interval within two to three categories of comparative net health benefit, then you have "low" certainty.

4. **Assign a joint rating in the Evidence Rating Matrix.** The final step is the assignment of the joint rating of magnitude of comparative net health benefit and level of certainty. As shown again in the figure on the following page, when your certainty is **"high,"** the estimate of net benefit is relatively assured, and so there are distinct labels available: a rating of **A** indicates a high certainty of a substantial comparative net benefit. As the magnitude of comparative net health benefit decreases, the rating moves accordingly, to **B** (incremental), **C** (comparable), and finally **D**, indicating an inferior or negative comparative net health benefit for the therapeutic agent relative to the comparator.

   When the level of certainty in the point estimate is only **"moderate,"** the summary ratings differ based on the location of the point estimate and the ends of the boundaries of the conceptual confidence interval for comparative net health benefit. The ratings associated with moderate certainty include **B+** (incremental or better), which indicates a point estimate of small *or* substantial net health benefit and a conceptual confidence interval whose lower end does not extend into the comparable range. The rating **C+** (comparable or better) reflects a point estimate of either comparable, small, *or* substantial net health benefit and a lower bound of the conceptual confidence interval that does not extend into the inferior range. These ratings may be particularly useful for new drugs that have been tested using noninferiority trial designs, or those involving modifications to an existing agent to provide adherence or safety advantages.

   Another summary rating reflecting moderate certainty is **P/I** (promising but inconclusive). This rating is used to describe an agent with evidence suggesting that it provides a comparable, small, or substantial net benefit over the comparator. However, in contrast to ratings **B+** and **C+, P/I** is the rating given when the conceptual confidence interval includes a small likelihood that the comparative net health benefit might actually be negative. In our experience the **P/I** rating is a common rating when assessing the evidence on novel agents that have received regulatory approval

with evidence of some benefit over placebo or the standard of care, but without robust evidence regarding safety profiles when used in community practice.

The final rating category is **I** (insufficient). This is used in two situations: (a) when there is moderate certainty that the best point estimate of a drug's comparative net health benefit is comparable, but there is judged to be a moderate-high likelihood that further evidence could reveal that the comparative net health benefit is actually negative; and (b) *any* situation in which the level of certainty in the evidence is **"low,"** indicating that limitations in the body of evidence are so serious that no firm point estimate can be given and/or the conceptual confidence interval for comparative net health benefit extends across all 4 categories. This rating would be a common outcome for assessments of the comparative effectiveness of two active drugs, when there are rarely good head-to-head data available; this rating might also commonly reflect the evidence available to judge the comparative effectiveness of a drug being used for an off-label indication.